Implementation with Near-Complete Information *

Kim-Sau Chung† Jeffrey C. Ely‡

July 29, 2002

Abstract

Many refinements of Nash equilibrium yield solution correspondences which do not have closed graph in the space of payoffs or information. This has significance for implementation theory, especially under complete information. If a planner is concerned that all equilibria of his mechanism yield a desired outcome, and entertains the possibility that players may have even the slightest uncertainty about payoffs, then the planner should insist on a solution concept with closed graph. We show that this requirement entails substantial restrictions on the set of implementable social choice rules. In particular, when preferences are strict (or more generally, hedonic), while almost any social choice function can be implemented in undominated Nash equilibrium, only monotonic social choice functions can be implemented in the closure of the undominated Nash correspondence.

*We benefited from conversations with Nabil Al-Najjar and Stephen Morris. Thanks to Eddie Dekel and Mike Whinston for helpful comments and advice. The suggestions of two anonymous referees helped us to improve the exposition of this paper.

†sau@nwu.edu

‡ely@nwu.edu. Financial Support from NSF grant SES 99-85462 is gratefully acknowledged.

1 Introduction

Results from the theory of implementation under complete information suggest that a planner can implement virtually any social choice function if he can expect that players will play according to some refinement of Nash equilibrium, say undominated Nash equilibrium or subgame-perfect Nash equilibrium. The mechanisms used in the proofs of these results make conspicuous use of the assumption of complete information, for example by severely punishing players when their reports are inconsistent, or by asking players to challenge one another to deter lying.

Complete information entails common knowledge of preferences, an assumption generally taken to be at best a simplifying approximation, but often less innocuous. In this paper we address the following question. Suppose the planner acknowledges that complete information is an idealization and that in the true environment players may be uncertain about the state of the world. What social choice functions can be implemented by mechanisms which provide the desired outcomes in all equilibria of environments that are arbitrarily close to complete information?

We find that this type of robustness analysis yields surprisingly strong restrictions. We demonstrate this by considering implementation in undominated Nash equilibrium (UNE-implementation). As shown by Palfrey and Srivastava (1991), almost any social choice function can be UNE-implemented in complete information environments. Without imposing any restrictions on the mechanism used, we find that only monotonic social choice functions can be robustly implemented when players have strict, or more generally, hedonic preferences.

The technical observation that lies at the heart of our conclusion is that refinements like undominated Nash equilibrium yield solution correspondences that do not have closed graph with respect to information, and discontinuities often occur at points of complete information. This means that even when *all* complete-information solutions yield the planner's desired outcomes, there may be environments arbitrarily close to complete information with equilibrium outcomes far from the desired set. In this paper we show that this is *necessarily* the case for mechanisms which implement non-monotonic social choice functions.

Related observations have been made in the game theory literature, for example in Fudenberg, Kreps, and Levine (1988), Dekel and Fudenberg (1990), and Kajii and Morris (1997). The result in the present paper, while similar in spirit, requires a distinct type of argument. The aforementioned papers study elaborations of complete-information games in which "crazy types" with altogether different preferences appear, and consequently, preferences over all possible action profiles are perturbed. In this paper, we fix the set of types from the complete-information environment and perturb only the information structure. Thus not all payoffs in the strategic form can be affected by the perturbation. Messages remain cheap talk.

2 Illustration of the Theorem

In this section we illustrate the logic of our main result within the context of a simple example. The example is taken from Jackson and Srivastava (1996, Example 5)[1] There are two players, labeled player 1 and player 2, two states of the world, labeled θ' and θ'', and three social alternatives a, b, and c among which a planner must choose. Players' strict preference orderings over these alternatives depend on the state of the world. Player 2 has preference $a \succ_2' b \succ_2' c$ in state θ', and preference $a \succ_2'' c \succ_2'' b$ in state θ''. Player 1 has the same ranking $c \succ_1 a \succ_1 b$ in either state. The planner designs a mechanism to implement a social choice function f which selects an alternative $f(\theta)$ for each state.

For the Nash equilibrium solution concept with complete information, Maskin (1977) derived a necessary condition, known as monotonicity, for implementability of a social choice function. (For a formal definition of monotonicity, see section 5.) On the other hand, when the planner uses a refinement of Nash equilibrium such as Nash equilibrium in undominated strategies (UNE), monotonicity is no longer necessary and in fact the set of implementable social choice functions can be quite large. Consider the non-monotonic social choice function f defined by $f(\theta') = a$ and $f(\theta'') = c$. The following mechanism Γ, in which player 1 chooses the row and player 2 chooses the column, implements f in UNE.

	m_2'	m_2''
m_1'	a	a
m_1''	b	c

The profile m', leading to outcome $a = f(\theta')$ is the unique Nash equilibrium in state θ', but both m' and m'' are Nash equilibria in state θ''. However, m_2'

[1]We thank an anonymous referee for suggesting this example.

being dominated for player 2 in state θ'', the only undominated Nash equilibrium is m'' leading to the outcome $c = f(\theta'')$.

The profile m' is ruled out by UNE because player 2 knows the true state with certainty. If player 2 is only nearly certain that the the true state is θ'', and hence entertains some (perhaps vanishingly small) probability that the true state is θ' (i.e., he actually prefers b to c), then the argument that m_2' is dominated no longer goes through.

In fact, in a sequence of vanishingly small perturbations of player 2's information, m_2' is never dominated. To see this observe that if player 1 were to play m_1' in state θ'' and m_2' in state θ', then m_2' would be a *strict* best-response for player 2 for any belief which assigns positive probability to state θ'. The implication is that m' would remain as an UNE along such a sequence of perturbations despite the fact that it is dominated at the complete-information limit. Formally, it means that the UNE correspondence induced by the mechanism Γ does not have closed graph, and due to the non-monotonicity of f there is a crucial discontinuity at the point of complete information.

However, to establish monotonicity as a necessary condition for robust UNE-implementation, we must show in that *any* mechanism that implements a non-monotonic social choice function such as f necessarily fails this robustness test. We show this in Theorem 1.

Let us conclude this section with a preview of some of the secondary aspects of the analysis to follow. First, in the type of perturbation used in the proof of our main theorem, some players are necessarily imperfectly informed about their own preferences. Indeed, for any perturbation of the example above this must be so. It is this type of perturbation that we exploit in the proof of Theorem 1. In many of the economic contexts in which implementation theory is applied, potential uncertainty of one's own preferences is a relevant concern. One prominent example arises in the application to the theory of contracts where the "planner's" uncertainty is about the result of actions previously taken by the players themselves. In section 5.2 , we present a simple version of this type of application and show the constraints that our robustness concept implies.

On the other hand, the full strength of our robustness test is less appropriate in truly *private-value* environments; i.e., situations in which the only relevant un-modeled uncertainty concerns the preferences of other players. Implementation of allocations in pure exchange economies would be a typical example. For these environments we formalize in Section 5.1 a weaker robustness test by considering only those near-complete information structures in which players' knowledge of their own preferences is preserved. Our Proposition 1 shows that within this class of perturbations, robustness of UNE-implementation is much less of a concern.

Finally, we remark on our approach to modeling players' preferences under incomplete information. To define equilibrium in incomplete information environments we must extend the state-dependent preference relation given by the original implementation problem to a preference relation over such uncertain prospects. Rather than assuming some specific form for such preferences, we employ just the axioms on preferences that we need for our results.

3 The Environment

There is a finite set N of players, and a set A of social alternatives. There is a finite set Θ of states of the world. Associated with each state θ is a preference profile $\succeq^\theta$ which is a list $(\succeq_1^\theta, \ldots, \succeq_n^\theta)$ where $\succeq_i^\theta$ is player i's state θ preference relation over A.

For our necessary conditions for robust implementability, we will assume players have *hedonic* preferences. Intuitively speaking, a player is said to have hedonic preferences if he does not care about aspects of social alternatives that only affect other players, but has strict preferences over aspects that affect he himself.

Definition 1 *Players are said to have hedonic preferences if for each player i, for every pair of states θ and θ', and for every pair of social alternatives a and b,*

$$a \sim_i^\theta b \Leftrightarrow a \sim_i^{\theta'} b.$$

Hedonic preferences are often assumed in, for example, the literature of matching and assignment problems.[2] For another example, suppose that A consists of various production levels of a public good, as well as taxation levels for each player. A given player will have strict preferences over levels of the public good and his own tax, but will be indifferent in every state over the tax applied to other players. Clearly the stronger assumption of *strict* preferences is a special case.

Players do not observe the state directly, but are informed of the state via

[2]See, for example, Pápai (2000).

signals. Player i's signal set is S_i which for simplicity we identify with Θ.[3] A signal profile is an element $s = (s_1, \ldots, s_n) \in S = \times_{i \in N} S_i$. When the realized signal profile is s, each player i observes only his own signal s_i. We let μ denote the prior probability over $\Theta \times S$, and let $\mathcal{P}$ be the set of all such priors.

Let us designate s^θ to be the signal profile in which each player's signal is the state θ. Complete information refers to the environment in which $\mu(\theta, s) = 0$ whenever $s \neq s^\theta$. Under complete information, the state, and hence the full profile of preferences is always common knowledge.

A social choice rule is a function $f : \Theta \to A$. A mechanism is a game form $\Gamma = (M, g)$. Here $M = \times_i M_i$ refers to the set of message profiles m, where $m = (m_1, \ldots, m_n)$. The outcome function $g : M \to A$ assigns to each message profile m an alternative $g(m) \in A$. Given a prior μ, a mechanism determines a Bayesian game $\Gamma(\mu)$ in which each player's type is his signal, and after observing his signal, player i selects a message from the set M_i. A strategy in $\Gamma(\mu)$ for player i is a rule $\sigma_i : S_i \to M_i$. A strategy profile $\sigma = (\sigma_1, \ldots, \sigma_n)$ lists a strategy for each player.

Under a complete information prior μ, the game $\Gamma(\mu)$ can be thought of as a collection of distinct strategic form games which can effectively be solved independently of one another because the preferences are common knowledge.

Definition 2 *Let μ be a complete information prior. We say that a strategy profile σ is a Nash equilibrium of $\Gamma(\mu)$ if for each player i, state θ, and message m_i,*

$$g(\sigma(s^\theta)) \succeq_i^\theta g(m_i, \sigma_{-i}(s^\theta)).$$

Let μ^θ denote the prior which assigns probability 1 to (θ, s^θ). In such a case it will not cause confusion to refer to a strategy σ_i simply by the action m_i it prescribes for signal $s_i = \theta$.

When the environment is not one of complete information (in which case we will say there is incomplete information) a player may not know the state, and hence the preference profile, with certainty. Moreover, since players other than i may be conditioning their behavior in the mechanism on information not available to player i (i.e., their signals), even in a given state, the outcome itself may appear random to player i. In order to describe the player's decision problem in such cases we must make an assumption about their preferences under

[3]This merely ensures that there are enough signals so that each player can be fully informed of the state. This allows us to accommodate complete information which is the focus of this paper. Nothing would be added by allowing for additional signals.

uncertainty. Rather than working with any specific model of such preferences, such as Bayesian expected utility maximization, we will assume only what will be needed for our result.

An act is a mapping $\alpha : \Theta \times S \to A$. Let $\mathcal{A}$ be the set of all acts. A belief is a probability β on $\Theta \times S$. The notation $C(\beta)$ denotes the support of β. We assume that for any given belief β each player i has a preference relation $\succeq_i^\beta$ over acts. We assume only the following about this family of preference relations:[4]

Assumption 1 *Let α and $\hat{\alpha}$ be two acts, and β a belief. Then*

$$[\alpha(\theta, s) \succeq_i^\theta \hat{\alpha}(\theta, s) \quad \forall(\theta, s) \in C(\beta)] \implies \alpha \succeq_i^\beta \hat{\alpha};$$

and if one of the preferences on the left-hand-side is strict for a state which has positive probability under β, then the preference on the right-hand-side is strict.

Obviously this axiom is implied by expected utility maximization as well as many other commonly studied theories of choice under uncertainty. Given an environment, and a model of preferences under uncertainty consistent with the above axiom, we can define the analog of Nash equilibrium for mechanisms under incomplete information. Let σ be a strategy profile in mechanism $\Gamma = (M, g)$. The act α_σ^Γ induced by σ is defined by $\alpha_\sigma^\Gamma(\theta, s) = g(\sigma(s))$. We furthermore assume that players derive conditional beliefs $\beta = \mu(\cdot|s_i)$ from the prior μ using Bayes' rule.

Definition 3 *Given a mechanism Γ, a Bayesian Nash equilibrium of $\Gamma(\mu)$ is a profile σ, such that for each player i, signal s_i, and strategy σ_i',*

$$\alpha_\sigma^\Gamma \succeq_i^{\mu(\cdot|s_i)} \alpha_{\sigma_i', \sigma_{-i}}^\Gamma.$$

In this paper, we will study implementation in undominated (Bayesian) Nash equilibrium (UNE); i.e.; (Bayesian) Nash equilibrium in which no player uses a dominated strategy. The following is a definition of *interim* weak-dominance for our setting.

Definition 4 *Let Γ be a mechanism. Strategy σ_i is dominated for type s_i if there exists a strategy σ_i' such that for every strategy profile σ_{-i} of players other than i, $\alpha_{\sigma_i', \sigma_{-i}}^\Gamma \succeq_i^{\mu(\cdot|s_i)} \alpha_\sigma^\Gamma$ with a strict preference for at least one σ_{-i}. Strategy σ_i is undominated if it is not dominated for any type.*

[4]Throughout, we fix a particular family $\{\succeq_i^\beta\}_\beta$ and our definitions of equilibrium and dominance are stated in terms of this family.

It is obvious that under a complete-information prior μ, a strategy profile σ is a Bayesian Nash equilibrium of $\Gamma(\mu)$ if and only if $\sigma(\theta)$ is a Nash equilibrium of $\Gamma(\mu^\theta)$ for each state. The same statement holds for undominated BNE and undominated NE.

4 $\overline{\text{UNE}}$-implementation

Henceforth we assume that A is a Hausdorff topological space, and that $\mathcal{A} = A^{\Theta\times S}$ is endowed with the product topology. Given a mechanism Γ, any solution concept $\mathcal{E}$ (such as Bayesian Nash equilibrium) induces a correspondence $\psi_\Gamma^\mathcal{E} : \mathcal{P} \to \mathcal{A}$, where each element α of $\psi_\Gamma^\mathcal{E}(\mu)$ is an act (or outcome) corresponding to some $\mathcal{E}$-solution of $\Gamma(\mu)$, which describes the alternatives $\alpha(\theta, s)$ that will result for each (θ, s).

Definition 5 *A mechanism Γ $\mathcal{E}$-implements a social choice function $f : \Theta \to A$ under μ if $\psi_\Gamma^\mathcal{E}(\mu) \neq \emptyset$, and for each $\alpha \in \psi_\Gamma^\mathcal{E}(\mu)$, we have $\alpha(\theta, s) = f(\theta)$ for each $(\theta, s) \in C(\mu)$.*

When μ is a complete-information prior, and the solution concept is Nash equilibrium (NE) or undominated Nash equilibrium (UNE), the above definition is equivalent to the standard definition of implementation.

Lemma 1 *Let μ be a complete information prior. Mechanism Γ NE-implements (resp. UNE-implements) a social choice function f if and only if for each state θ such that $\mu(\theta, s^\theta) > 0$, $\psi_\Gamma^{\mathrm{NE}}(\mu^\theta) \neq \emptyset$ (resp. $\psi_\Gamma^{\mathrm{UNE}}(\mu^\theta) \neq \emptyset$), and for each $m^* \in \psi_\Gamma^{\mathrm{NE}}(\mu^\theta)$ (resp. $m^* \in \psi_\Gamma^{\mathrm{UNE}}(\mu^\theta)$), $g(m^*) = f(\theta)$.*

When the solution correspondence $\psi_\Gamma^\mathcal{E}$ does not have closed graph, there may be environments arbitrarily close to μ where the set of solutions is undesirably large. Such a scenario would undermine the planner's confidence in his mechanism if he entertained the possibility that he had even slightly misspecified the environment. This motivates us to consider the "closure" of the solution correspondence $\psi_\Gamma^\mathcal{E}$. Define

$$\overline{\psi_\Gamma^\mathcal{E}}(\mu) = \{\alpha : (\mu, \alpha) \in \overline{\operatorname{graph} \psi_\Gamma^\mathcal{E}}\}.$$

The following notation will be convenient. If $\mathcal{B}$ is a set of acts such that $\alpha(\theta, s) = f(\theta)$ for each $\alpha \in \mathcal{B}$ and $(\theta, s) \in C(\mu)$, then we will write $\mathcal{B} \sqsubset_\mu f$.

Definition 6 *A mechanism* Γ $\overline{\mathcal{E}}$*-implements a social choice function* f *under* μ *if*

1. $\psi^{\mathcal{E}}_{\Gamma}(\mu) \neq \emptyset$,
2. $\overline{\psi^{\mathcal{E}}_{\Gamma}}(\mu) \sqsubset_{\mu} f$.

In this paper, we study implementation in the closure of the undominated Nash equilibrium correspondence, which we refer to as $\overline{\text{UNE}}$-implementation. In particular, we are interested in complete-information environments.

5 Monotonicity as a Necessary Condition

Recall the definition of monotonicity (Maskin (1977)):

Definition 7 *A social choice function* f *is monotonic if for every pair of states* θ *and* θ' *such that for each player* i,

$$a \succ_i^{\theta'} f(\theta) \implies a \succ_i^{\theta} f(\theta), \tag{1}$$

we have $f(\theta') = f(\theta)$.

Theorem 1 *Assume players have hedonic preferences. Then if a social choice function is* $\overline{\text{UNE}}$*-implementable under complete information, it is necessarily monotonic.*

Remark: To see that this is a substantial restriction, note that with strict preferences, it follows from Palfrey and Srivastava (1991) that with at least three players any social choice function that satisfies no-veto-power is UNE-implementable. Furthermore, hedonic preferences are consistent with the class of "separable" environments studied by Jackson, Palfrey, and Srivastava (1994) where *any* social choice function is implementable via a finite mechanism, even in the two-player case. We have not imposed any restrictions on the mechanism, such as boundedness (Jackson (1992)) or ruling out integer games (Sjöström (1994)).

Proof: Let complete-information prior μ be given, and let f be a $\overline{\text{UNE}}$-implementable social choice function with implementing mechanism $\Gamma = (M, g)$. Suppose θ and θ' are two possible states satisfying (1).

Since Γ $\overline{\text{UNE}}$-implements f, it also UNE-implements f since UNE $\subset$ $\overline{\text{UNE}}$. Thus, there exists an undominated Nash equilibrium m^* of $\Gamma(\mu^\theta)$ such that $g(m^*) = f(\theta)$. From (1) it follows that m^* is a Nash equilibrium of $\Gamma(\mu^{\theta'})$ for if not, there must exist a player i and a message m_i such that $g(m_i, m^*_{-i}) \succ_i^{\theta'} g(m^*)$. But by (1), this implies that $g(m_i, m^*_{-i}) \succ_i^{\theta} g(m^*)$, which is a contradiction since m^* is a Nash equilibrium $\Gamma(\mu^\theta)$.

If m^* is undominated in $\Gamma(\mu^{\theta'})$, then $m^* \in \psi_\Gamma^{\text{UNE}}(\mu^{\theta'})$, and since Γ UNE-implements f, it follows from Lemma 1 that $f(\theta') = f(\theta)$, and we are done. So suppose m^* is dominated in $\Gamma(\mu^{\theta'})$. Then let $I \subset N$ be the non-empty set of players i for whom m_i^* is dominated in $\Gamma(\mu^{\theta'})$; and for each $i \in I$, let D_i be the set of dominating messages. Because m^* is undominated in $\Gamma(\mu^\theta)$, for each $m_i \in D_i$, one of the following must hold

1. $g(m_i^*, \hat{m}_{-i}) \sim_i^\theta g(m_i, \hat{m}_{-i})$ for all $\hat{m}_{-i}$.
2. There exists $\hat{m}_{-i}$ such that $g(m_i^*, \hat{m}_{-i}) \succ_i^\theta g(m_i, \hat{m}_{-i})$.

In case 1, by the assumption of hedonic preferences, $g(m_i^*, \hat{m}_{-i}) \sim_i^{\theta'} g(m_i, \hat{m}_{-i})$ for all $\hat{m}_{-i}$, but this contradicts the fact that m_i^* is dominated by m_i for i in $\Gamma(\mu^{\theta'})$. Hence, we must have case 2 for each $m_i \in D_i$.

Consider the following family of information structures ν^ε, parameterized by $\varepsilon > 0$. Let τ^i represent the profile of signals $(s_1, \ldots, s_n)$ defined by $s_i = \theta'$ and $s_j = \theta$ for all $j \neq i$.

$$\nu^\varepsilon(\theta, \tau^i) = \frac{\varepsilon}{|I|}\mu(\theta, s^\theta), \qquad \forall i \in I, \tag{2}$$

$$\nu^\varepsilon(\theta, s^\theta) = (1-\varepsilon)\mu(\theta, s^\theta), \tag{3}$$

$$\nu^\varepsilon(\tilde{\theta}, s^{\tilde{\theta}}) = \mu(\tilde{\theta}, s^{\tilde{\theta}}), \qquad \forall \tilde{\theta} \neq \theta. \tag{4}$$

In this information structure, when the state is anything other that θ or θ', the state is common knowledge. Furthermore, when a player observes the signal θ, that player knows that the state is θ. Obviously $\nu^\varepsilon \to \mu$ as $\varepsilon \to 0$. Note that

$$C(\nu^\varepsilon) = \{(\tilde{\theta}, s^{\tilde{\theta}}) : \tilde{\theta} \in \Theta\} \cup \{(\theta, \tau^i) : i \in I\}.$$

Let $\hat{\sigma}$ be an undominated Nash equilibrium of the complete information game $\Gamma(\mu)$. (We know there is at least one because Γ UNE-implements f.) We consider the strategy profile σ in which $\sigma_i(s_i) = \hat{\sigma}_i(s_i)$ for $s_i \notin \{\theta, \theta'\}$, and $\sigma_i(\theta) = \sigma_i(\theta') = m_i^*$. We claim that for every $\varepsilon > 0$, this profile is an undominated Bayesian Nash equilibrium of $\Gamma(\nu^\varepsilon)$. This will finish the proof because

σ generates an act α^Γ_σ for which $\alpha^\Gamma_\sigma(\theta', s^{\theta'}) = f(\theta)$ and since σ is an undominated BNE for every $\varepsilon > 0$, it follows that $(\mu, \alpha^\Gamma_\sigma) \in \overline{\text{graph}\,\psi^{\text{UNE}}_\Gamma}$. Thus, since Γ $\overline{\text{UNE}}$-implements f, we must have $f(\theta') = \alpha^\Gamma_\sigma(\theta', s^{\theta'}) = f(\theta)$.

Consider any σ'_i for player i. The act generated by σ'_i against σ_{-i} is given by $\alpha^\Gamma_{\sigma'_i,\sigma_{-i}}(\theta, s) = g(\sigma'_i(s_i), \sigma_{-i}(s_{-i}))$. By construction, for every $(\tilde{\theta}, s) \in C(\nu^\varepsilon)$, the message profile played is a Nash equilibrium of $\Gamma(\mu^{\tilde{\theta}})$, and hence $g(\sigma(s)) \succeq^{\tilde{\theta}}_i g(\sigma'_i(s_i), \sigma_{-i}(s_{-i}))$. Thus Assumption 1 implies $\alpha^\Gamma_\sigma \succeq^{\nu^\varepsilon(\cdot|s_i)}_i \alpha^\Gamma_{\sigma'_i,\sigma_{-i}}$ for each s_i. This establishes that σ is a BNE of $\Gamma(\nu^\varepsilon)$.

We now show that σ_i is undominated for each i. By construction, for each type $s_i \notin \{\theta, \theta'\}$, $\sigma_i(s_i)$ is undominated for type s_i because under ν^ε, for each such type, the preference profile is common knowledge, and σ_i selects an undominated Nash equilibrium for these preferences.

A player of type $s_i = \theta$ knows under ν^ε that the true state is θ, and hence that his preferences are $\succeq^\theta_i$. However, it does not follow immediately that $\sigma_i(s_i) = m^*_i$ is undominated for s_i, because in the environment ν^ε, type s_i assigns positive conditional probability to more than one type profile. If i's opponents use strategies that play different messages in these different type profiles, then from the perspective of player i, they are playing a mixed (possibly even correlated) strategy profile. We must ensure that m^*_i is undominated for i against such possibly mixed strategy profiles.

Since m^*_i is undominated in μ^θ, we know that for every message m_i, either case 1 or 2 is satisfied. In case 1, it follows from Assumption 1 that for every strategy profile σ'_{-i} of the opponents of i, $\alpha^\Gamma_{m^*_i,\sigma'_{-i}} \sim^{\nu^\varepsilon(\cdot|s_i)}_i \alpha^\Gamma_{m_i,\sigma'_{-i}}$, because these acts are pointwise indifferent from the point of view of type $s_i = \theta$. In case 2, we can set $\sigma'_j(\cdot) \equiv \hat{m}_j$ for all $j \neq i$, and it again follows from Assumption 1 that $\alpha^\Gamma_{m^*_i,\sigma'_{-i}} \succ^{\nu^\varepsilon(\cdot|s_i)}_i \alpha^\Gamma_{m_i,\sigma'_{-i}}$. Together, these imply that no m_i can dominate m^*_i for type $s_i = \theta$.

Finally, consider type $s_i = \theta'$. If $i \notin I$, then under ν^ε, such a type knows that his preferences are $\succeq^{\theta'}$. Since m^*_i was not dominated in $\Gamma(\mu^{\theta'})$, it follows from an argument analogous to the one in the previous paragraph that m^*_i is not dominated for type $s_i = \theta'$ under ν^ε.

Suppose on the other hand, $i \in I$. For any $m_i \in D_i$, there exists $\hat{m}_{-i}$ satisfying case 2. In this case we set $\sigma'_j(s_j = \theta) = \hat{m}_j$ and $\sigma'_j(s_j = \theta') = m^*_j$ for all $j \neq i$. Against this strategy profile, the act resulting from i playing m^*_i gives $g(m^*_i, \hat{m}_{-i})$ in state θ and $g(m^*)$ in state θ'. Since m^* is a Nash equilibrium under $\mu^{\theta'}$, $g(m^*) \succeq^{\theta'}_i g(m_i, m^*_{-i})$ and by construction $g(m^*_i, \hat{m}_{-i}) \succ^\theta_i g(m_i, \hat{m}_{-i})$.

Thus, Assumption 1 implies $\alpha^{\Gamma}_{m_i^*,\sigma'_{-i}} \succ_i^{\nu^{\varepsilon}(\cdot|s_i)} \alpha^{\Gamma}_{m_i,\sigma'_{-i}}$, and hence m_i does not dominate m_i^* for type $s_i = \theta'$.

The last case to consider is $m_i \notin D_i$. If case 2 is satisfied for m_i, then by an identical argument, type $s_i = \theta'$ strictly prefers to play m_i^* against the σ'_{-i} constructed according to the previous paragraph. On the other hand, if case 1 is satisfied, then $g(m_i^*, \hat{m}_{-i}) \sim_i^{\nu^{\varepsilon}(\cdot|s_i)} g(m_i, \hat{m}_{-i})$ for every $\hat{m}_{-i}$. Thus, m_i does not dominate m_i^* for type $s_i = \theta'$. ■

5.1 "Private" Values

In the proof of Theorem 1 we construct a near-complete information structure in which there is asymmetric information about the state of the world. We use the fact that some players have superior information about the state and consequently the preferences of other players.

In this section we restrict attention to *private-value* information structures: priors ν which satisfy $\nu(\theta, s) = 0$ whenever $\succeq_i^{\theta} \neq \succeq_i^{s_i}$ (recall that we identify S_i with Θ). Thus while type s_i may be uncertain about the state and hence the types and preferences of the other players, he knows that his own preferences are $\succeq_i^{s_i}$. Let $\tilde{\mathcal{P}}$ be the set of such private-value priors. Given a mechanism, any strategy which is dominated under complete information is also dominated under any $\nu \in \tilde{\mathcal{P}}$.

Proposition 1 *If σ'_i dominates σ_i for type s_i under complete information, then σ'_i dominates σ_i for type s_i under any private value prior $\nu \in \tilde{\mathcal{P}}$.*

Proof: Let θ be the state identified with signal s_i. We have

$$g(\sigma'_i(s_i), m_{-i}) \succeq_i^{\theta} g(\sigma_i(s_i), m_{-i}) \tag{5}$$

for every profile of messages m_{-i}, with strict preference for some $\hat{m}_{-i}$. Fix $\nu \in \tilde{\mathcal{P}}$. We claim that σ'_i dominates σ_i under ν for type s_i. Indeed (5) holds for every m_{-i} with θ replaced by any $\theta' \in C(\nu(\cdot|s_i))$ because the private values assumption implies $\succeq_i^{\theta'} = \succeq_i^{\theta}$. It now follows from Assumption 1 that for any strategy profile σ_{-i},

$$\alpha^{\Gamma}_{\sigma'_i,\sigma_{-i}} \succeq_i^{\nu(\cdot|s_i)} \alpha^{\Gamma}_{\sigma_i,\sigma_{-i}},$$

with strict preference when $\sigma_{-i}(\cdot) \equiv \hat{m}_{-i}$. ■

It follows that any Nash equilibrium that is dominated when information is complete will remain dominated after any perturbation of the information structure in which players' knowledge of their own preferences is preserved.[5]

5.2 An Application

Consider a standard principal-agent problem. At date 0, the principal and the agent sign a contract. At date 1, the agent exerts either high or low effort, which translates into the principal's future profitability y. At date 2, the principal and the agent play any message game described in the date-0 contract, and payments are made according to the result of the message game. We assume away any renegotiation of the contract, so that if there is problem for implementation it is not due to renegotiation.

Assume for simplicity that the translation from the agent's effort to the principal's future profitability y is deterministic and one-to-one, and hence $y \in \{y_L, y_H\}$ with $y_L < y_H$. If y is both observable and verifiable, then the principal can implement the first-best effort level with a contract that conditions payments on y. Let f^* be a first-best contract. Note that f^* must be deterministic if both players are risk averse.

Suppose y is observable but not verifiable. Then y would become the "state" at date 2 when the principal and the agent play any message game described in the date-0 contract. The date-0 contract design problem is equivalent to our mechanism design problem with $\Theta = \{y_L, y_H\}$ and A equal to some subset of lotteries over $\{(t_P, t_A) \in \mathbf{R}^2 : t_P + t_A \leq 0\}$, where t_P and t_A are payments to the principal and the agent respectively. The implementation problem is whether or not f^* can be implemented in some solution concept.

In the standard principal-agent problem, it is usually assumed that the agent has the same preferences over A once his date-1 effort is sunk. If the principal is risk neutral or has an exponential utility function, then she too will have the same preferences over A regardless of the state y, and hence the lack of preference reversal would imply that the non-constant f^* is not implementable. So let's assume that the principal's preferences exhibit decreasing absolute risk aversion: $-u''(y + t_P)/u'(y + t_P)$ is strictly decreasing in $(y + t_P)$.

[5] It does not follow from Proposition 1 that UNE-implementation alone is sufficient for UNE-implementation. As discussed in Duggan and Roberts (1997), if the original implementing mechanism is badly behaved, the Nash correspondence itself may not have closed graph.

If A is rich in the sense that we can always punish one player without changing the lottery of payments received by the other player, then we are in the "separable" environment of Jackson, Palfrey, and Srivastava (1994), and hence by their Theorem 3, f^* is UNE-implementable.

However, since it is the agent's effort that determines the principal's future profitability y, it is likely that the principal does not have full confidence on her observation of y, and suspects that the agent knows slightly better than she does about the true state (and hence her own preferences). So it is important to ask whether or not f^* is also $\overline{\text{UNE}}$-implementable. Suppose, on top of being "separable," A is furthermore discrete. Then generically each player will have strict preferences over the possible lotteries he or she may receive. Since it is natural for each player not to care about the lottery received by the other player, players' preferences are hedonic. Hence, by Theorem 1, only monotonic social choice functions can be $\overline{\text{UNE}}$-implementable.[6] But then f^*, which is both non-constant and deterministic, cannot possibly be $\overline{\text{UNE}}$-implementable, because any lottery that is better than $f^*(y_H)$ at state y_L must also be better than $f^*(y_H)$ at state y_H, and hence monotonicity would have required that $f^*(y_L) = f^*(y_H)$.

5.3 Sufficiency

Our robustness test delivers monotonicity as a necessary condition for implementability. Of course, this would be a trivial result if robustness were in fact so strong as to, say, render little more than dictatorial social choice functions implementable. So for completeness, we demonstrate that a slight strengthening of Maskin's sufficient conditions for Nash implementability implies $\overline{\text{UNE}}$-implementability. We however caution the reader that our implementing mechanism, like those in the literature on Nash implementation, uses an integer game.

To prove our result we will need to strengthen our assumptions on players' preferences under uncertainty. To Assumption 1 we add a continuity property:

Assumption 2 *For every pair of acts α and $\hat{\alpha}$, the set $\{\beta : \alpha \succ_i^{\beta} \hat{\alpha}\}$ is open.*

We first strengthen Maskin's monotonicity condition:

[6] In fact, because only the principal's preferences are state-dependent, just as in the example in Section 2, it is enough to perturb only the principal's information to obtain monotonicity as a necessary condition.

Definition 8 *A social choice function f is strongly monotonic if for every pair of states θ and θ' such that for each player i,*

$$a \succ_i^{\theta'} f(\theta) \implies a \succeq_i^{\theta} f(\theta),$$

we have $f(\theta') = f(\theta)$.

Strong monotonicity is actually equivalent to monotonicity in many economic applications. For example, consider the economic environment where there exists a private good that is both desirable and continuously transferable. Consider a monotonic social choice function f. If $f(\theta) \neq f(\theta')$, then there exist a player i and an alternative a such that $a \succ_i^{\theta'} f(\theta)$ and yet $f(\theta) \succeq_i^{\theta} a$. But then there exists an alternative b which is the same as a except that player i receives slightly less of the private good. By continuity of preferences we will have $b \succ_i^{\theta'} f(\theta)$ and $f(\theta) \succ_i^{\theta} b$. Hence f is strongly monotonic as well. Strong monotonicity and monotonicity will also be equivalent when players have hedonic preferences.

We shall also strengthen Maskin's no-veto-power condition:

Definition 9 *Let Y be a subset of alternatives. A social choice function f satisfies Y-no-veto-power if whenever there is an alternative $a \in A$ such that for at least $N-1$ players i, $a \succeq_i^{\theta} b$ for every $b \in Y$, we have $f(\theta) = a$.*

We will assume Y-no-veto-power for a *finite* set Y because this will enable us to construct an implementing mechanism which has finite range. The equilibria of this mechanism will be strict, and the associated *finite* set of strict inequalities can be preserved by a small enough perturbation. Finite Y-no-veto-power is equivalent to standard no-veto-power in environments in which for each player and state there is an alternative that is best for that player in that state. Simply take Y to be the (finite) set of alternatives that are best for some player at some state. The two versions of no-veto-power will also be equivalent in economic environments where alternatives that are very good for one player are necessarily bad for all other players.

Theorem 2 *Suppose there are at least 3 players. If f is strongly monotonic, satisfies Y-no-veto-power for some finite $Y \subset A$, and if for each player i and state θ there is an alternative $a(i,\theta)$ such that $f(\theta) \succ_i^{\theta} a(i,\theta)$, then f is $\overline{\text{UNE}}$-implementable.*

Proof: We construct an implementing mechanism $\Gamma = (M, g)$. For each i, the set of available messages is $M_i = \Theta \cup (\mathbb{Z} \times Y)$ where $\mathbb{Z}$ is the set of integers. That is, each player is asked to report *either*[7] a state, or an integer and an alternative. The outcome function g is defined as follows. Let m^θ denote the message profile $(\theta, \theta, \ldots, \theta)$, and $m^\theta \setminus m_i$ the profile obtained from m^θ by substituting m_i for player i. We set:

- $g(m^\theta) = f(\theta)$;
- if m_i is a state θ', and if there exists an alternative a such that $a \succ_i^{\theta'} f(\theta)$ but $f(\theta) \succ_i^{\theta} a$, then $g(m^\theta \setminus m_i) = a$ (if there is more than one such a, select one arbitrarily);
- if m_i is not a state, then $g(m^\theta \setminus m_i) = a(i, \theta)$;
- if m contains at least three distinct reports, and if each m_i is a state, then $g(m)$ is an arbitrary element of $f(\Theta)$;
- if at least one player has announced an integer and an alternative, then $g(m)$ is the alternative named by the player who named the greatest integer (break ties in any deterministic way).

We prove that Γ $\overline{\text{UNE}}$-implements f in three steps. Let complete-information prior μ be given. We first show that there is a neighborhood U of μ such that $\Gamma(\nu)$ has an undominated BNE for every $\nu \in U$. Next, we demonstrate that *every* Nash equilibrium of $\Gamma(\mu)$ yields f, i.e. $\psi_\Gamma^{\text{NE}}(\mu) \sqsubset_\mu f$. Finally, we show that $\overline{\psi_\Gamma^{\text{NE}}}(\mu) \subset \psi_\Gamma^{\text{NE}}(\mu)$. Since $\overline{\psi_\Gamma^{\text{UNE}}}(\mu) \subset \overline{\psi_\Gamma^{\text{NE}}}(\mu)$, this proves the result.

Step 1: Consider the truthful strategy profile $\sigma_i(\theta) = \theta$ for each i. It yields outcome $g(m^\theta) = f(\theta)$. By construction, if in state θ, player i sends message $m_i \neq \theta$, the outcome $g(m^\theta \setminus m_i)$ is strictly worse for i according to $\succeq_i^\theta$. Hence, σ_i is a *strict* Nash equilibrium under complete information. We now show that for every strict Nash equilibrium σ under complete information, there is a neighborhood U of μ in which the equilibrium remains a strict Bayesian Nash equilibrium.

Define

$$\mathcal{B}_i = \{\alpha^\Gamma_{\sigma_i', \sigma_{-i}} : \sigma_i' \neq \sigma_i\}$$

[7] This option, together with the finiteness of Y are what distinguish this mechanism from the usual one.

Note that although i has infinitely many strategies, the set $\mathcal{B}_i$ of acts that are available to i against σ_{-i} is finite. Since σ is a strict Nash equilibrium in $\Gamma(\mu)$, we have $\alpha^{\Gamma}_{\sigma} \succ_i^{\mu(\cdot|s_i)} \alpha$ for each s_i and $\alpha \in \mathcal{B}_i$. By Assumption 2, and the finiteness of N, S, and $\mathcal{B}_i$'s, there is a neighborhood U of μ such that for every $\nu \in U$, we have $\alpha^{\Gamma}_{\sigma} \succ_i^{\nu(\cdot|s_i)} \alpha^{\Gamma}_{\sigma'_i,\sigma_{-i}}$ for each i, s_i, and $\sigma'_i \neq \sigma_i$. Thus, σ remains a strict and hence undominated Bayesian Nash equilibrium for each $\Gamma(\nu)$.

Step 2: Suppose σ is a Nash equilibrium of $\Gamma(\mu)$. We will show that $\alpha^{\Gamma}_{\sigma} \sqsubset_{\mu} f$. First suppose that in $\sigma(\theta)$, each player announces the same state θ'. Then $g(\sigma(\theta)) = f(\theta')$. In this case we claim $f(\theta) = f(\theta')$, otherwise by strong monotonicity there exists a player i and an alternative a such that $a \succ_i^{\theta} f(\theta')$ but $f(\theta') \succ_i^{\theta'} a$, and in this case we have constructed Γ so that $g(\sigma(\theta) \setminus \theta)$ is some such a. Thus $\sigma(\theta)$ would not be a Nash equilibrium of $\Gamma(\mu^{\theta})$. For any other profile $\sigma(\theta)$, there must be at least $N-1$ players who can deviate from $\sigma(\theta)$ and bring about a profile in which there are at least 3 distinct messages. Thus, by construction of Γ, each of these players could dictatorially choose his θ-most-preferred alternative from Y. But since $\sigma(\theta)$ is a Nash equilibrium of $\Gamma(\mu^{\theta})$, it must be that for each of these players i, $g(\sigma(\theta)) \succeq_i^{\theta} a$ for every $a \in Y$. Since f satisfies Y-no-veto-power, $f(\theta) = g(\sigma(\theta))$ as desired.

Step 3: Consider any act α such that $\alpha(\theta, s) \neq f(\theta)$ for some $(\theta, s) \in C(\mu)$. To show that $\alpha \notin \overline{\psi^{\mathrm{NE}}_{\Gamma}}(\mu)$ it suffices to find a neighborhood U of μ such that $\alpha \notin \psi^{\mathrm{NE}}_{\Gamma}(\nu)$ for each $\nu \in U$. This is because the set of available acts in the mechanism Γ is finite and $\mathcal{A}$ is Hausdorff. Set $\Sigma^{\alpha} = \{\sigma : \alpha^{\Gamma}_{\sigma} = \alpha\}$. If Σ^{α} is empty, there is nothing to prove. So assume Σ^{α} is not empty. Define

$$\mathcal{W}^{\alpha}_i = \{\alpha^{\Gamma}_{\sigma'_i,\sigma_{-i}} : \sigma \in \Sigma^{\alpha} \text{ and } \exists s_i \text{ s.t. } \alpha^{\Gamma}_{\sigma'_i,\sigma_{-i}} \succ_i^{\mu(\cdot|s_i)} \alpha\}.$$

The sets $\mathcal{W}^{\alpha}_i$ are finite because the set of available acts in the mechanism Γ is finite. Hence, by Assumption 2, there is a neighborhood U of μ such that for every $\nu \in U$, $i \in N$, and $\hat{\alpha} \in \mathcal{W}^{\alpha}_i$, there exists s_i such that $\hat{\alpha} \succ_i^{\nu(\cdot|s_i)} \alpha$.

Consider any $\sigma \in \Sigma^{\alpha}$. By Step 2 σ cannot be a Nash equilibrium of $\Gamma(\mu)$. There is thus a player i and a strategy σ'_i such that $\alpha^{\Gamma}_{\sigma'_i,\sigma_{-i}} \in \mathcal{W}^{\alpha}_i$, and hence there exists a type s_i such that $\alpha^{\Gamma}_{\sigma'_i,\sigma_{-i}} \succ_i^{\nu(\cdot|s_i)} \alpha$ for each $\nu \in U$. It follows that σ is not a Bayesian Nash equilibrium of $\Gamma(\nu)$. Since σ was an arbitrary element of Σ^{α}, we have shown that $\alpha \notin \overline{\psi^{\mathrm{NE}}_{\Gamma}}(\mu)$.

To conclude the proof, we summarize:

$$\emptyset \neq \overline{\psi^{\mathrm{UNE}}_{\Gamma}}(\mu) \subset \overline{\psi^{\mathrm{NE}}_{\Gamma}}(\mu) \subset \psi^{\mathrm{NE}}_{\Gamma}(\mu) \sqsubset_{\mu} f.$$

■

References

Dekel, E., and D. Fudenberg (1990): "Rational Behavior with Payoff Uncertainty," *Journal of Economic Theory*, 52, 243–267.

Duggan, J., and J. Roberts (1997): "Robust Implementation," http://www.economics.utoronto.ca/roberts/Robust16.pdf.

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

Jackson, M. O. (1992): "Implementation in Undominated Strategies: A Look at Bounded Mechanisms," *Review of Economic Studies*, 59, 757–775.

Jackson, M. O., T. R. Palfrey, and S. Srivastava (1994): "Undominated Nash Implementation in Bounded Mechanisms," *Games and Economic Behavior*, 6, 474–501.

Jackson, M. O., and S. Srivastava (1996): "Characterizations of Game Theoretic Solution Concepts which Lead to Impossibility Theorems," *Review of Economic Studies*, 63(1), 23–38.

Kajii, A., and S. Morris (1997): "The Robustness of Equilibria to Incomplete Information," *Econometrica*, 65(6), 1283–1310.

Maskin, E. (1977): "Nash Equilibrium and Welfare Optimality," mimeo.

Palfrey, T. R., and S. Srivastava (1991): "Nash Implementation Using Undominated Strategies," *Econometirca*, 59, 479–501.

Pápai, S. (2000): "Strategyproof Assignment by Hierarchical Exchange," *Econometrica*, 68, 1403–1433.

Sjöström, T. (1994): "Implementation in Undominated Nash Equilibrium without Integer Games," *Games and Economic Behavior*, 6, 502–511.