Soc Choice Welfare (2001) 18: 655–708

Social Choice and Welfare

A crash course in implementation theory

Matthew O. Jackson

Humanities and Social Sciences 228-77, California Institute of Technology, Pasadena, CA 91125, USA (e-mail: jacksonm@hss.caltech.edu)

Received: 26 March 2001/Accepted: 21 May 2001

Abstract. This paper is meant to familiarize the audience with some of the fundamental results in the theory of implementation and provide a quick progression to some open questions in the literature.

1 Introduction

There are many economic, social, and political situations where individuals interact to make decisions that affect them collectively. Examples range from voting to elect representatives or choose a public policy, to trading in a market. Based on their preferences over the possible outcomes of the interaction, individuals may act strategically in order to influence the outcome to their advantage. For instance, in an election a voter might vote for his or her second ranked candidate if the voter's favorite candidate has no chance of being elected; or in an auction a buyer may select a bid considering tradeoffs between the probability of winning the auction and the price to be paid. The specific design of the institution through which individuals interact, for instance the rules of the election or auction, can have a profound impact on the strategic behavior of the members of the society and on the outcomes of the process. Implementation theory is a study of the relationship between the structure of the institution through which individuals interact and the outcome of that interaction.

This paper grew out of lectures given at the NATO Advanced Study Institute on Game Theory and Resource Allocation: The Axiomatic Approach, which took place at SUNY Stony Brook in July of 1997. I thank the organizers and especially William Thomson for organizing the institute and the participants for feedback on the lectures. I thank Salvador Barbera for detailed suggestions an earlier version of this manuscript, and Maurice Salles and an anonymous referee for helpful comments.

Game theory plays a central role in the modeling of the strategic interaction studied in implementation theory. In many applications of game theory, the game modeling interaction is taken as given and analyzed to predict the actions of individuals and the resulting outcome. In implementation theory, instead of taking the game as given, it is something to be designed. Often, one thinks of the desired outcomes as the given and analyses whether there exist game forms for which the strategic properties induce individuals to (always) choose actions that lead to the desired outcomes. An example of an implementation question is: how can we design an auction to be sure that the individual who most highly values an object is sure to be the winner of the auction?[1] In this view, implementation theory is a normative branch and game theory is a positive branch of the same tree, and implementation theory is the design or reverse engineering process associated with game theory. Of course, this view is a bit caricatured, but indicates that there is a close relationship between the tools and understandings developed in implementation theory and game theory.

To get a feeling for the type of questions that are analyzed in implementation theory let us start by looking at a classic example. Consider a society or committee holding an election to select one out of a set of candidates. Each member of the society has a preference ranking over the candidates. The society may have certain aspirations regarding which candidate should be selected as a function of the preferences of the members of the society. For instance, it may wish to avoid selecting a candidate who is Pareto dominated by another candidate (i.e., a candidate ranked lower than another candidate by all members of society). It may also wish to select a Condorcet winner (a candidate who defeats any other candidate in a pairwise comparison according to a majority of voters' preferences) if such a candidate exists. If these were the goals of the society then the implementation question would be, "Does there exist an election procedure for which for each possible profile of preference rankings of the voters, each equilibrium outcome of the election procedure would be Pareto optimal and Condorcet consistent?"

In order to answer this question one has to make precise what an election procedure is and what equilibrium outcomes are. This is the point at which game theoretic tools are used. The election procedure is modeled as a game form or what is commonly referred to as a mechanism in the literature. It specifies a set of possible actions or messages that each member of society can use, and then the outcome (in this case the candidate elected) as a function of the actions or messages sent by the members of society. For instance one could have each member of society submit their ranking of the candidates. If there are m candidates, then one could award a candidate m points for each voter whose submitted ranking places them highest, $m-1$ points for each voter's submitted ranking places them second highest, and so on.

[1] For an analysis of this particular problem from a mechanism design point of view, see Dasgupta and Maskin (1997). I will discuss the relationship and difference between mechanism design and implementation.

The elected candidate is the one who has the most points, with ties broken according to some pre-specified rule. This is the mechanism corresponding to Borda's (1781) scoring method. Of course, it may not be in a society member's best interest to report their true preference ranking. This is where equilibrium concepts from game theory are used to make predictions concerning strategic behavior. A solution concept such as Nash equilibrium can be used to predict the preference rankings that will be reported by voters as a function of the voters' true preference rankings. Depending on the class of mechanisms that are admitted (e.g., how complicated we allow the message spaces to be) and which equilibrium concept that is used, we can end up with different answers concerning the possibility of selecting Pareto optimal outcomes and being Condorcet consistent.[2]

In the above example a very concrete question was posed. We asked whether it was possible to design election procedures that possess specific properties, or in the language of the theory, whether it is possible to implement correspondences (mapping preference profiles into social outcomes) with specific properties. Also, we focused our attention on a specific setting. While this is one way to proceed in analyzing implementation, it does not *directly*[3] provide a general understanding of implementation that moves across different problems. Another way to proceed, is to take a more abstract approach to the implementation problem and attempt to fully characterize what can be implemented. That is, we can search for properties that precisely identify which correspondences are implementable and which are not, and also identify implementing mechanisms. These properties can then be applied to specific settings to check whether a given correspondence can be implemented. Such general characterization results have been produced with remarkably little in terms of assumptions on the structure of the set of alternatives or the feasible preferences of individuals. As we shall see, the theory has alternated between fairly general and abstract theorems that cover a wide variety of potential applications, and more focused and detailed theorems that apply to particular settings or correspondences.

In what follows, I concentrate on a specific progression in the literature and in doing so provide a biased view that often reflects my own perspective on the literature. These lectures are meant to be an introduction to the theory rather than a survey,[4] and so I do this without apology. At points I provide

[2] The Borda scoring mechanism will not satisfy Condorcet consistency. Moreover, as we shall see in Example 3, if we use Nash equilibrium as a solution concept then it will be impossible to always select Pareto optimal outcomes and be Condorcet consistent. However, as shown for instance in Dutta and Sen (1993) and Jackson et al. (1994), under the solution concepts of backwards induction or undominated Nash equilibrium it will be possible to always select Pareto optimal outcomes and be Condorcet consistent.

[3] Nevertheless, I will come back to discuss why an approach of cataloguing the answers to such narrower questions may turn out to be a useful approach to developing the theory.

[4] Surveys of various aspects of the literature may be found in Moulin (1982), Moore (1992), Palfrey (1992, 1995), Palfrey and Srivastava (1993), Allen (1997), and Corchon

opinions concerning assumptions and results, which are quite critical. This is meant to constructively point out limitations of some of the existing results and suggest potential directions for future research.

In what follows I assume that the reader has an introductory level knowledge of game theory (and thus has seen notions such as Nash equilibrium) and is also familiar with some concepts from microeconomics such as preference relations and Walrasian equilibrium.

2 Definitions and an example

2.1 Individuals

A finite group of *individuals* interact. N denotes both the set of individuals and its cardinality. Generic individuals are represented as i, j, and k.

2.2 Outcomes

The set of *outcomes* is denoted A, and generic elements are represented as a, b, c, d.

The set of outcomes may be finite or infinite depending on the application. For example, consider the design of a voting procedure to elect one of a finite number of candidates. In that case, N is the set of voters and A is the finite set of candidates. As another example, consider the design of a market where individuals interact to exchange ℓ different goods. In that case N is the set of economic agents, and $A \subset \mathbb{R}_+^{N\ell}$ represents the final allocation of goods (including labor, leisure, consumption goods, etc.) that are possible given the endowments and production possibilities.

2.3 Preferences

Individual i's *preferences* are represented by a binary relation R_i over A that is complete and transitive. The notation aR_ib indicates that i weakly prefers alternative a to b. The strict preference relation associated with R_i is denoted P_i (where aP_ib if and only if it is not the case that bR_ia). The notation R denotes a profile $R = (R_1, \ldots, R_N)$, and $(R_{-i}, \bar{R}_i)$ denotes the profile where the i-th entry of R is replaced with $\bar{R}_i$.

The set of admissible profiles of preferences is the set $\mathscr{P}$. Depending on the application, $\mathscr{P}$ may impose restrictions on the preferences. For instance, in the context of the exchange of private goods it may be assumed that preferences are convex, continuous, and non-decreasing. I will sometimes refer to a profile of preferences as being the state of the environment.

(1998). While these lectures are not meant to survey the literature, the Bibliography that I include here is fairly comprehensive. The exception is that I do not include references to the large literature related to dominant strategy implementation as Salvador Barberà (2001) offers lectures on strategy-proofness in this same volume.

2.4 Social choice correspondences

A *social choice correspondence*, F, maps profiles of preferences into subsets of alternatives. For any $R \in \mathscr{P}$, $F(R) \subset A$ represents the set of socially desirable alternatives when preferences are given by R. It will be assumed throughout that F is non-empty. When F is single-valued it is referred to as a *social choice function*.

In many applications, F will be a well-known correspondence, such as the Walrasian, Lindahl, or top-cycle correspondence, or will be a social choice correspondence derived from some normative axioms, such as the correspondence of Pareto optimal, individually rational, and envy-free allocations in a private good exchange setting.

Generally, implementation theory seeks to characterize the set of social choice correspondences that are obtainable as equilibrium outcomes when individuals interact through some game form making strategic use of their knowledge of the preference profile R. There are several perspectives that one can take on this problem. One perspective is to begin with a specific correspondence and setting, such as the Walrasian correspondence in a private-goods economic setting, and ask whether that specific correspondence can be implemented and if so by what mechanism. A second perspective is to work more abstractly and characterize the set of correspondences that are implementable, by identifying conditions that they must satisfy. Both of these (and some other) perspectives have been taken in the literature, and we shall see examples of each.

2.5 Mechanisms

A *mechanism* is a pair M, g, where $M = M_1 \times \cdots \times M_N$ is a cross product of message spaces and $g : M \to A$ is an outcome function. Thus, for each profile of messages $m = (m_1, \ldots, m_N)$, $g(m) \in A$ represents the resulting outcome or allocation.

A mechanism is often also referred to in the literature as a *game form*. The terminology game form distinguishes it from a game, as the consequence of a profile of messages (or actions) is an outcome rather than a vector of utility payoffs. Once the preferences of the individuals are specified, then a game form induces a game. Since in the implementation analysis the preferences of individuals vary from state to state, this distinction between game forms and games is critical.

2.6 Solution concepts

A *solution concept*[5] specifies the strategic behavior of individuals faced with a mechanism (M, g) given a preference profile R. Thus, it is a correspondence

[5] I use the terminology "solution concept" rather than "equilibrium concept", as some of the concepts employed in the literature (such as single or iterative removal of dominated strategies or maximin) do not make explicit use of "equilibrium" ideas, and can predict vectors of messages that are not stable in an equilibrium sense.

S that identifies a subset of M for any given (M, g, R) specification. For the question of what a mechanism implements, the specific messages that are predicted by the solution concept are only of intermediate interest as the corresponding set of outcomes that result is the important concept. Thus, we pay attention to the outcome correspondence associated with a solution S represented by $O_S(M, g, R) = \{a \in A \mid \exists m \in S(M, g, R) \text{ s.t. } g(m) = a\}$.

For example, for any given (M, g, R) a *pure strategy Nash equilibrium* is a profile $m \in M$ such that $g(m)R_i g(m_{-i}, \bar{m}_i)$ for all i and $\bar{m}_i \in M_i$. Denote this set by $NE(M, g, R)$. The set of associated outcomes is $O_{NE}(M, g, R) = \{a \in A \mid \exists m \in NE(M, g, R) \text{ s.t. } g(m) = a\}$.

I define other solution concepts as they arise and similarly defer definitions of mixed strategies until they arise.

2.7 Implementation

A social choice correspondence F is *implemented by the mechanism* (M, g) via the solution S, if $O_S(M, g, R) = F(R)$ for all $R \in \mathscr{P}$. F is said to be *implementable* via the solution S if there exists a mechanism (M, g) which implements it.

The above form of implementation is sometimes referred to as *full* implementation as it requires the exact coincidence of the outcomes of a mechanism with the social choice correspondence.

One can also reasonably argue for a weaker form of implementation where one only requires that $O_S(M, g, R)$ be a non-empty subset of $F(R)$ for every R.[6] This takes the point of view that any outcome in $F(R)$ is socially desirable and so any selection is fine. Note, however, that F is weakly implementable if and only if some sub-correspondence of F is implementable. The literature has thus proceeded in characterizing what is implementable, which then provides an indirect answer to the question of weak implementability.[7]

One key aspect of implementation (and weak implementation) is the requirement that *all* equilibrium outcomes lie in the given social choice correspondence. This is different from designing a mechanism that has one desired outcome as an equilibrium outcome, as there may also be undesired equilibrium outcomes that are not accounted for. This points to the important distinction between the "implementation" literature and the "mechanism design" literature as being one of worrying about multiple equilibria.[8] The mechanism design literature focuses on incentive compatibility issues, asking whether a given outcome can be induced as *an* equilibrium of some mechanism and generally ignores whether there are other equilibria. In situations where this leads to a negative result (e.g., Myerson and Satterthwaite 1983) showing that certain outcomes cannot be sustained as the equilibrium of any mechanism, this

[6] See Thomson (1996) for an argument supporting full implementation over weak.

[7] This may not be an entirely satisfactory approach, as there may be conditions for weak implementability that are easier to verify than checking whether some sub-correspondence satisfies the conditions for implementability.

[8] See Jackson (2000) for a recent survey of the mechanism design literature.

approach is fine. However, in situations where it leads to positive results so that one identifies a mechanism and one of its equilibrium as satisfying some desired criteria, then one should be very careful in interpreting the result. Can we be sure that the given equilibrium will be played and not some other equilibrium of the mechanism?[9] Are there other mechanisms that are better from the perspective of maintaining the desired equilibrium, but not having other undesired ones? These questions are important and can be applied to almost all of the mechanism design literature, including the auctions and the principal-agent literatures.[10] This point is well illustrated in an example of Demski and Sappington (1984), where the "optimal" mechanism to a principal-multiple agent problem has a second equilibrium which makes all the agents better off and the principal worse off.[11]

The multiple equilibrium issue is why the implementation problem is still non-trivial in environments with complete information, where each individual knows the other individuals' preferences. As an example, suppose that we consider trying to implement the Walrasian correspondence in a classical Edgeworth setting. If we only want to make sure that Walrasian outcomes are Nash equilibrium outcomes, and do not care about what other equilibrium outcomes may arise, there is a simple mechanism that works. Individuals simultaneously announce a full vector of feasible allocations. If they all announce the same vector of allocations then that is the outcome, and otherwise the outcome is that they each keep their endowment. Clearly everyone announcing the same Walrasian allocation is a Nash equilibrium of this mechanism. Unfortunately, there are many other equilibria of this mechanism which are not Walrasian allocations. The implementation question is whether there exists a mechanism whose entire set of Nash equilibrium outcomes coincides with the set of Walrasian allocations for each preference profile.

2.8 On the timing and information structure

Much of the implementation literature uses an idiom that refers to a *social planner*, who is benevolent and selects the mechanism to implement a social choice correspondence with a society's best interest at heart. A question that then naturally comes to mind when thinking about the implementation problem is why doesn't the social planner adjust the mechanism once the preference profile is realized.[12] Alternatively, why do the individuals in society have

[9] If so, why was this not built into the solution concept to begin with?

[10] For further examples and a pointed discussion of this issue, see Palfrey and Srivastava (1993).

[11] A defense that is often offered for focusing a single equilibrium is that society can direct its members to play the desired equilibrium and then they should have no reason to deviate. This argument loses some power when there are alternative equilibria that are preferred by some (or even all) of the members of society, especially if there is any possibility for the members to directly or indirectly coordinate on an equilibrium.

[12] See Baliga et al. (1997) for the possibility of having the planner be a player who has preferences and can affect the outcome function as part of equilibrium play.

complete information concerning each others' preferences (or at least enough information to arrive at an equilibrium) and the social planner not? An answer is that mechanisms represent institutions (or in some cases even constitutions), that are meant to be long-lived. These may be costly to adjust, or it even may be better to commit to fixing the institutions not to adjust to the specifics of each realized setting. The rules for a securities market are fixed and then traders arrive with their desired trades, and similarly the rules for an election are (usually) set before preferences of the voters are realized. Thus, mechanisms can be thought of as representing institutions, and the implementation problem is to characterize institutions in terms of the correspondence of outcomes they produce as a function of the realized state of individual preferences.

Finally, while incomplete information environments, where for instance individuals know their own preferences and hold beliefs over the preferences of others, are arguably more realistic than complete information environments, it has proven useful to attack the problem one piece at a time. Understanding implementation in the complete information setting has helped significantly in developing characterizations of implementation in Bayesian settings.

2.9 An example.

Hurwicz laid much of the foundation of the implementation problem, and presented a simple example that is a very good one for a first analysis of some of the issues.

Example 1 (Hurwicz 1972). Consider a two person, two good exchange economy, represented by an Edgeworth box. Individual 1 has an endowment $e^1 = (0,1)$ and individual 2 has an endowment $e^2 = (1,0)$. Let x^i_ℓ denote individual i's allocation of the ℓ-th good. So here our space of feasible allocations is $A = \{(x^1, x^2) \in \mathbb{R}^4_+ \,|\, x^1_1 + x^2_1 = x^1_2 + x^2_2 = 1\}$.

Consider two states of the world. In the first state (R^1, R^2), the preference relation R^i of each individual is represented by a Cobb-Douglas utility function $u_i(x) = x^i_1 x^i_2$. In the second state, $(\bar{R}^1, R^2)$, the preferences of agent 1, $\bar{R}^1$, are represented by the utility function $\bar{u}_1(x) = x^1_2 - \dfrac{1}{1 + x^1_1}$.

Suppose that we are interested in implementing the Walrasian correspondence. The unique Walrasian equilibrium allocation at R^1, R^2 is $x^1 = x^2 = (1/2, 1/2)$, and at $\bar{R}^1, R^2$ is $x^1 = (1/2, 7/9)$ and $x^2 = (1/2, 2/9)$.

Clearly, in either state agent 1 prefers to have the allocation associated with $(\bar{R}^1, R^2)$ to that associated with (R^1, R^2).[13] This is seen in Fig. 1.

There are a few important conclusions to draw from this example. The first is that a direct mechanism, or the classic one associated with a "Walrasian auctioneer" who asks individuals to report their demand functions, will not implement the Walrasian correspondence. Individual 1 is better off acting as

[13] Hurwicz actually makes a stronger point. Every allocation that is Pareto efficient and individually rational for both individuals at $\bar{R}^1, R^2$, is preferred under R^1 (and $\bar{R}^1$) to the Walrasian equilibrium allocation at R^1, R^2.

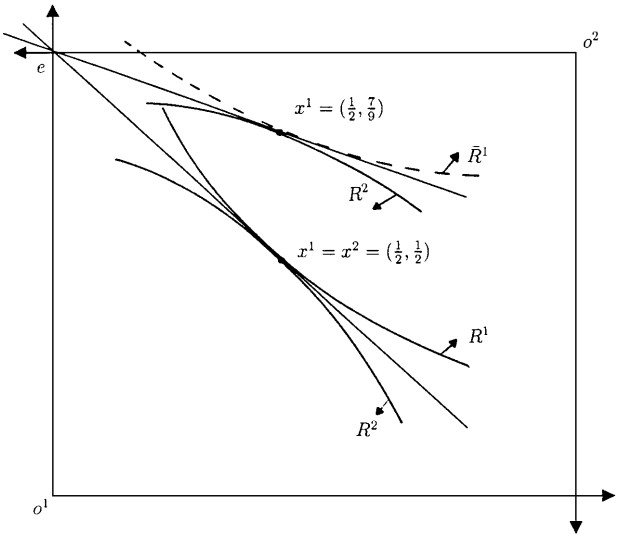

Fig. 1.

if he had preferences $\bar{R}^1$ instead of R^1 when the state is actually (R^1, R^2). This also implies that any sort of incentive compatibility condition will be violated by the Walrasian correspondence (or any selection from it) when agents' preferences are private information.

Hurwicz points out another consequence of this line of reasoning. A mechanism that implements[14] the Walrasian correspondence *must* make use of information of agent 2 concerning the preferences of agent 1, as agent 1 can otherwise pretend to be of type $\bar{R}^1$. Thus, Hurwicz argues that no "privacy preserving" mechanism can implement the Walrasian correspondence, where privacy preserving is an appropriate formalization of the idea that the mechanism only incorporates knowledge of an individual about his or her own preferences.

Let us now explore a related question in the context of the same problem in order to get a better feeling for the implementation problem. We know that a mechanism that implements the Walrasian correspondence in this example must make some use of individual 2's knowledge of individual 1's preferences. So we might ask, is that enough? Is there a simple mechanism that Nash implements the Walrasian correspondence in this two state example?

The following is the direct mechanism associated with simply asking individuals to announce their own preferences. The first entry in each row is individual 1's allocation and the second is individual 2's allocation.

Clearly, there is a unique strictly dominant strategy equilibrium in either state for individual 1, which is to say $\bar{R}^1$, and so such a direct mechanism

[14] I am being vague about the solution concept at the moment, but the statement is true for most solutions concepts.

Table 1.

		Individual 2 R^2
Individual 1	R^1	$x^1 = (\frac{1}{2}, \frac{1}{2})$, $x^2 = (\frac{1}{2}, \frac{1}{2})$
	$\bar{R}^1$	$x^1 = (\frac{1}{2}, \frac{7}{9})$, $x^2 = (\frac{1}{2}, \frac{2}{9})$

Table 2.

		Individual 2	
		Left	Right
Individual 1	R^1	$x^1 = (\frac{1}{2}, \frac{1}{2})$, $x^2 = (\frac{1}{2}, \frac{1}{2})$	$x^1 = (\frac{11}{18}, \frac{2}{3})$, $x^2 = (\frac{7}{18}, \frac{1}{3})$
	$\bar{R}^1$	$x^1 = (0, 1)$, $x^2 = (1, 0)$	$x^1 = (\frac{1}{2}, \frac{7}{9})$, $x^2 = (\frac{1}{2}, \frac{2}{9})$

cannot Nash implement the Walrasian correspondence. This is just an illustration of Hurwicz's point.[15]

Let us now expand individual 2's strategies to include two possible announcements "Left" and "Right". We will design the mechanism so that it is a strictly dominant strategy for individual 1 to say R^1 when the state is R^1, and $\bar{R}^1$ when it is $\bar{R}^1$. Moreover, this can be done so that it is a strict best response for individual 2 to say "Left" if individual 1 is playing R^1, and to say "Right" if individual 1 is saying $\bar{R}^1$.

For the mechanism in Table 2, the unique Nash equilibrium in state (R^1, R^2) is (R^1, Left) which results in the Walrasian allocation for that state, and similarly for the state $(\bar{R}^1, R^2)$ the unique Nash equilibrium is $(\bar{R}^1, \text{Right})$. This mechanism thus implements the Walrasian correspondence (for this simple two state environment) in Nash equilibrium. The mechanism also implements the desired correspondence in the iterative elimination of strictly dominated strategies, and most other solution concepts.

As is consistent with Hurwicz's privacy preserving point, it was necessary for us to have individual 2 play a role even though only 1's preferences vary. One thing to note about the above mechanism, is that individual 2 does not simply "report" what she knows about individual 1. Individual 2 would have an incentive to say that it was state R^1, even if it was state $\bar{R}^1$. Thus, the agents' incentives in this regard are opposed. Instead, allowing individual 2 to have a non-trivial message space allows us to create a richer mechanism; and in particular to exploit switches in individual 1's preferences between the two states. For instance, the "$\bar{R}^1$, Left" entry is one that individual 1 prefers to "R^1, Left"

[15] This also illustrates that no mechanism can implement the Walrasian correspondence in dominant strategies, as it would have to take this form when looking only at the strategies used based on these preferences.

in state $\bar{R}^1$ but not in state R^1. Thus, the off-diagonal entries are critical to the working of the mechanism and were not chosen by chance. This begins to hint at an important necessary condition for Nash implementation that I discuss next.

The contrast between the workings of the two mechanisms above points out the importance of considering a large class of mechanisms for the implementation problem, and most notably to include mechanisms that do more than simply ask individuals to report their preferences. Although the mechanism in Table 2 can be regarded as a direct mechanism if one thinks of asking each individual to report the state, the entries in the off-diagonals are still critical to successful implementation. Moreover, there are problems where implementation can only be achieved by mechanisms that are more complicated. In order to develop a deeper understanding of this, let us now analyze Nash implementation.

3 Nash implementation

The seminal work on Nash implementation by Maskin (1999),[16] not only provides us with an understanding of what is implementable in Nash equilibrium, but it also provides a blueprint for the techniques and approach that underlie many of the general characterization results in the literature. Thus, it is useful to study Nash implementation in some detail.

3.1 Monotonicity two ways

Maskin identified an intuitive necessary condition for Nash implementation that he called monotonicity.[17] This condition may be expressed in two different ways. It is a trivial exercise to see that the statements are equivalent. Nevertheless, I present each of them separately as these expressions correspond to different expressions in the implementation problem, and it is useful to emphasize each of these points of view.

Suppose that a social choice correspondence F is Nash implementable. What can we deduce about F? Its implementability implies that there exists an implementing mechanism (M, g). So consider a preference profile R and an alternative $a \in F(R)$. Since F is Nash implementable there exists a profile of actions $m \in M$ such that m is a Nash equilibrium at R and $g(m) = a$. Next, suppose that there exists another preference profile $\bar{R}$ such that $a \notin F(\bar{R})$. The fact that (M, g) Nash implements F then implies that m cannot be a Nash equilibrium at $\bar{R}$. Thus, there must exist an agent i and a deviation $\bar{m}_i$ that i prefers to m_i at $\bar{R}_i$. That is, there must exist i and $\bar{m}_i$ such

[16] The paper circulated as a working paper from 1977 to 1998.

[17] This condition previously appeared in the social choice literature under the name strong positive association (e.g., see Muller and Satterthwaite 1977). The name monotonicity, however, stuck and has recently been referred to as Maskin-monotonicity by several authors. This condition differs from the monotonicity condition discussed in the context of social welfare orderings (e.g., see Moulin 1988 for a review of some results regarding monotonic social welfare orderings).

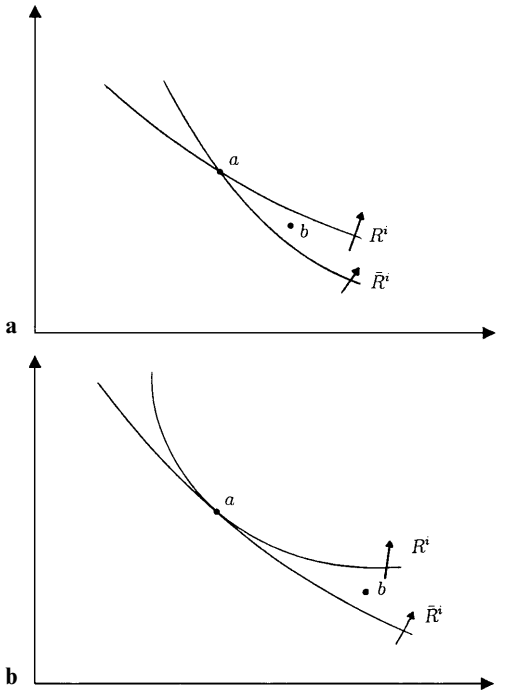

Fig. 2a,b.

that $g(m_{-i}, \bar{m}_i)\bar{P}_i g(m)$. Since m was a Nash equilibrium at R, it must be that $g(m)R_i g(m_{-i}, \bar{m}_i)$. Letting $b = g(m_{-i}, \bar{m}_i)$, we have reasoned that F must satisfy the following condition.

A social choice correspondence F is *monotonic* if for any $R \in \mathscr{P}$, $\bar{R} \in \mathscr{P}$, and $a \in F(R)$ such that $a \notin F(\bar{R})$, there exist i and b such that $aR_i b$ and $b\bar{P}_i a$.

This condition is represented in Fig. 2, above.

In Fig. 2, the set of allocations is in $\mathbb{R}^2_+$ for individual i. The monotonicity condition states that if $a \in F(R)$ but $a \notin F(\bar{R})$, then there exists some i for whom either the indifference curves through a corresponding to R_i and $\bar{R}_i$ cross (Fig. 2a)[18] or the upper contour set of $\bar{R}_i$ through a is a superset of the upper contour set of R_i through a (Fig. 2b).

We have reasoned the following result.

Theorem 1 (Maskin 1999). *If a social choice correspondence F is Nash implementable, then F is monotonic.*

To get some practice understanding monotonicity, the reader can return to the implementing mechanism in Example 1 (Table 2) and verify that player 1

[18] Although in the figure, the indifference curves cross at a, they can cross at any point and may cross more than once. As noted in Maskin (1999), the Spence-Mirrlees single-crossing condition is stronger than monotonicity. For a study of the relationship between monotonicity and single crossing properties see Arya et al. (1998).

and the off-diagonal entries satisfy the monotonicity conditions that are necessary for implementation.

To develop a fuller understanding of monotonicity, let us examine the condition from a different, but formally equivalent perspective.

Consider a preference profile R and alternative $a \in F(R)$. Suppose that $\hat{R}$ is such that for each i if aR_ib, then $a\hat{R}_ib$. This says that a's standing in i's preference ranking has has not fallen from R to $\hat{R}$, so that a is still weakly preferred under $\hat{R}_i$ to each b it was weakly preferred to under R_i. From Nash implementability we know that there exists m which is a Nash equilibrium at R. This implies that $g(m)R_ig(m_{-i}, \hat{m}_i)$ for any i and potential $\hat{m}_i$. Given that aR_ib implies $a\hat{R}_ib$, it must be that $g(m)\hat{R}_ig(m_{-i}, \hat{m}_i)$ for any i and potential $\hat{m}_i$. Thus, m is also a Nash equilibrium at $\hat{R}$ and so $a \in F(\hat{R})$. This leads to the following statement of monotonicity.

A social choice correspondence F is *monotonic* if for any R, $a \in F(R)$, and $\hat{R}$, such that for each i and b aR_ib implies $a\hat{R}_ib$, $a \in F(\hat{R})$.

Thus, monotonicity requires that if $a \in F(R)$ and for each i the upper contour set of $\hat{R}_i$ through a is a subset of the upper contour set of R_i through a, then $a \in F(\hat{R})$.

This condition is represented in Fig. 3, below.

One statement of monotonicity follows the reasoning that if an alternative is to be implemented at one profile but not another, then it must have fallen in someone's rankings in order to break the Nash equilibrium via some deviation. The other statement of monotonicity follows the reasoning that if an alternative is implemented at one profile and rises in each individual's rankings at another preference profile, then profile of actions leading to the alternative which form a Nash equilibrium at the first profile must still be a Nash equilibrium profile at the second profile. These conditions are obviously equivalent as one is simply the contra-positive of the other. Nevertheless, it is still useful to consider both statements. The first emphasizes that there must exist some preference reversal if an equilibrium at one profile is to be broken at another. This suggests natural relationships with various preference crossing properties, such as the Spence-Mirrlees condition. The second emphasizes that if the

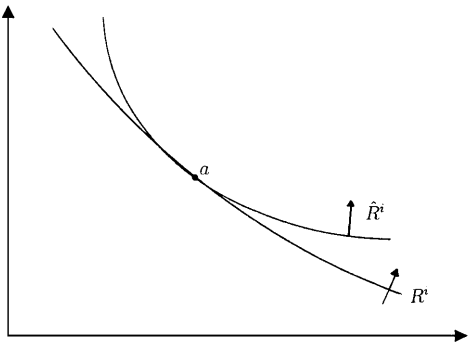

Fig. 3.

standing of an equilibrium alternative improves it must remain an equilibrium outcome, which is often a useful way of checking whether monotonicity is satisfied and provides useful bridges to understanding the connection to conditions such as strategy-proofness.

3.2 Sufficient conditions for Nash implementation

Monotonicity alone is not a sufficient condition for Nash implementation, but it is sufficient together with an auxiliary condition called no-veto power when there are at least three individuals.

A social choice correspondence F satisfies *no veto power* if whenever i, R, and a are such that aR_jb for all $j \neq i$ and all $b \in A$, then $a \in F(R)$.

No veto power is a condition that is quite weak in some contexts, such as environments with some private goods when there are at least three (non-satiated) individuals. In such a context, each individual would prefer to have more of the private good and so there is never a single alternative that is most preferred by more than one individual. However, the condition is more restrictive with 2 individuals, as then it requires that each individual's favorite outcome be in the correspondence; and in contexts such as voting environments where there is no private good, as then it is possible to have a number of individuals agree on a favorite outcome and no-veto power ignores how poorly that outcome may rank under a remaining individual's preferences.

Theorem 2 (Maskin 1999). *If $N \geq 3$ a social choice correspondence F satisfies monotonicity and no veto power, then it is Nash implementable.*

It is instructive to sketch a proof of this theorem, as it provides insight into the approach used to prove many of the sufficiency theorems in the literature.[19]

Generally, it is easy to design a mechanism that has the desired outcomes as Nash equilibria. It is more difficult to rule out undesired outcomes. To see this, consider the following trivial mechanism: if at least $N-1$ individuals name the same alternative then that is the outcome, otherwise some arbitrary b is the outcome. Notice that a unanimous announcement of any a is a Nash equilibrium for any preference profile. However, every alternative is an equilibrium outcome for such a mechanism regardless of the preference profile. This points to the importance of the multiple equilibrium problem. One way to view the mechanism below is to start with such a trivial mechanism and then modify it using monotonicity to rule out undesired equilibria.

Consider a social choice correspondence F which satisfies monotonicity and no veto power and assume that $N \geq 3$. Consider the following mechanism.

[19] Although the theorem was first stated in Maskin (1999) (circulating as a working paper in 1977), the first proof was by Williams (1986). William's proof required some additional assumptions on the setting, and the first complete proof of the theorem as stated is in Saijo (1988). The proof given here is adapted from Repullo (1987) and Moore and Repullo (1990). Karl Vind was the first to suggest this structure of proof, and a mechanism that was a precursor to this one, in a discussion of Maskin's work in the 1970's (see Groves 1977, Footnote 6).

$M_i = \mathscr{P} \times A \times \mathbb{N}$ where $\mathbb{N}$ is the set of nonnegative integers. Thus, each individual announces a preference profile, an alternative, and an integer.

Define g as follows:

(1) If $m_1 = m_2 = \cdots = m_N = (R, a, n)$ and $a \in F(R)$, then $g(m) = a$.
(2) If there exists i such that $m_j = (R, a, n)$ for all $j \neq i$, where $a \in F(R)$, and $m_i = (\cdot\,, b, \cdot)$ where $m_i \neq m_j$, then $g(m) = b$ if aR_ib and $g(m) = a$ if bP_ia.
(3) For any other m label $m_i = (\cdot, a_i, n_i)$ and let i^* be the lowest indexed i such that $n_i \geq n_j$ for all $j \neq i$, and then $g(m) = a_{i^*}$.

To paraphrase the mechanism:

(1) applies if all individuals make exactly the same announcement, and then the announced alternative a is the outcome of the mechanism.

(2) applies to situations where if all but one individual make exactly the same announcement, and some individual i makes a different announcement. In that case, the outcome is the b announced by that individual i if it is no better than the alternative a announced by the others under the preference for *i announced by the other individuals*. Otherwise the outcome is a.

(3) applies the remaining announcements that do not fall under (1) or (2). Here we identify the individual who announced the highest integer (and break ties according to individual labels) and the outcome is the alternative announced by that individual.

Let us verify that this mechanism Nash implements the desired F. Let R be the true preference profile.

First, let us check that for any $a \in F(R)$ it is a Nash equilibrium for all agents to announce $m_i = (R, a, 0)$. A unilateral deviation by any i results in an $m_{-i}, \bar{m}_i$ that falls in (2), and then can only result in a b such that aR_ib. Thus, no i can gain by deviating and so m is an equilibrium.

Next, let us check that every Nash equilibrium results in some $a \in F(R)$. If m is a Nash equilibrium that falls into (3), then it must be that $g(m)$ is the most preferred outcome of all agents, since otherwise any agent i could deviate and announce an integer higher than the other agents and select any outcome a_i. Thus, by no veto power it must be that $g(m) \in F(R)$. If m is a Nash equilibrium that falls into (2), then any agent $j \neq i$ (where i is as defined in (2)) could unilaterally deviate and choose an action such that g would be determined by (3) and $j = i^*$ (simply by announcing an integer higher than any other agent and any desired a_j). Thus, for such an m to be an equilibrium, it must be that $g(m)$ is most preferred by all agents $j \neq i$, and again no veto power ensures that $g(m) \in F(R)$. Finally, consider the case where m is a Nash equilibrium and g is determined by (1). If the preference profile R is announced truthfully, then the outcome must be in $F(R)$. So consider the case where $m_1 = m_2 = \cdots = m_N = (\tilde{R}, a, n)$ and $a \in F(\tilde{R})$ where $\tilde{R} \neq R$. Any individual i could deviate to obtain any $b \neq a$ such that $a\tilde{R}_ib$ (by simply announcing $(\cdot\,, b, \cdot)$ which puts $m_{-i}, \tilde{m}_i$ under (2)). Thus, it must be that if $a\tilde{R}_ib$ then aR_ib. Thus, by monotonicity $a \in F(R)$.

This basic structure of this mechanism underlies the construction of mechanisms in many of the constructive proofs for other solution concepts too. The

basic idea is that there is a possibility for complete agreement which falls under (1). This allows outcomes in $F(R)$ to be sustained as equilibrium outcomes. As it is possible for the individuals all to announce the same (R, a, n) even when the state is $\bar{R}$ and $a \notin F(\bar{R})$, part (2) of the definition of g allows monotonicity to work so that some i can deviate and announce $(\bar{R}, b, n)$ and obtain b. Part (3) then allows any individual other than i to obtain any outcome they like, which rules out equilibria that fall under (2), unless b is most preferred by all $j \neq i$ in which case by no veto power, $b \in F(\bar{R})$.

As discussed previously, no veto power is not a necessary condition and can be restrictive. So it is worthwhile to identify a set of conditions that is both necessary and sufficient for Nash implementation. Such conditions have been obtained by Moore and Repullo (1990). Although these conditions are more complicated, they do provide additional insight and I refer the reader to the Moore and Repullo (1990) paper for more detail.[20]

Theorem 2 requires that $N \geq 3$. The case of $N = 2$ is of obvious importance as there are many bilateral interactions that one would like the theory to handle. Interestingly, there are non-trivial differences between the case of $N = 2$ and $N \geq 3$. To get a rough idea of the additional considerations that appear for the case of $N = 2$, notice that the mechanism used to prove Theorem 2 makes use of the fact that when $N - 1$ agents announce the same thing, then one can identify an i who is demanding a different alternative as in (2) of the definition of g. With 2 agents this is not possible and an additional necessary condition appears, which arises from considerations of what happens when agent 1 plays a strategy that should be played in one state and agent 2 plays a strategy that should be played in another state. Characterizations for the case of $N = 2$ appear in Dutta and Sen (1991b) and Moore and Repullo (1990). While the full characterization is complex, an intuitive sufficient condition called the "non-empty lower intersection condition" appears in Dutta and Sen (1991b), and I refer the interested reader there for details.

3.3 The restrictiveness of monotonicity

There are aspects of monotonicity that make it a strong and restrictive condition, and a condition that is more easily motivated as a necessary condition of Nash implementation than by direct normative argument. Consider $a \in F(R)$ and $\hat{R}$ such that for each i and b aR_ib implies $a\hat{R}_ib$. Monotonicity requires that $a \in F(\hat{R})$. Two aspects of the strength of this condition are as follows. First, it is allowed for a to be strictly preferred to b under R_i and indifferent to b under $\hat{R}_i$. Thus, there is a limited sense in which a could have gotten worse in moving from R to $\hat{R}$, and yet monotonicity still requires that a lie in $F(\hat{R})$. The reasoning based on Nash implementation makes it clear as to why this is a necessary condition, but this makes it a condition that will be

[20] See Danilov (1989) and Yamato (1992) for elegant full characterizations in some specific settings and Sjöström (1991) for an algorithm for verifying the necessary and sufficient conditions.

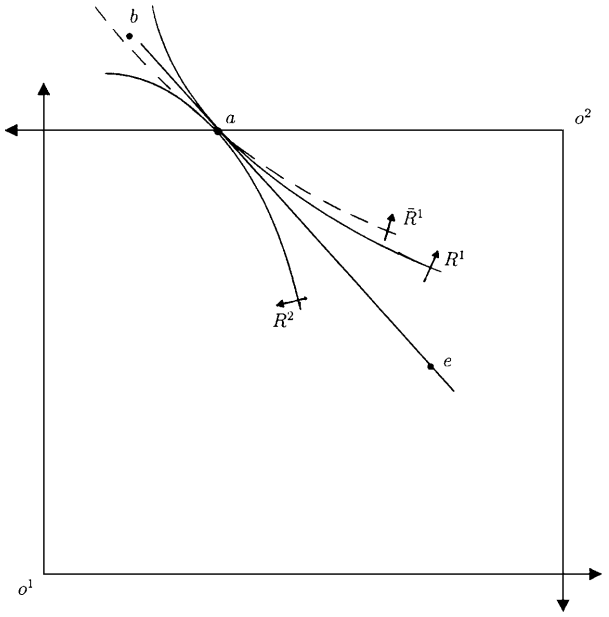

Fig. 4.

violated by a number of social choice correspondences that have some strong normative backing. Second, the comparison between R and $\hat{R}$ pays no attention to what happens to the relative ranking of alternatives other than a, (except those implied by comparisons to a). These aspects mean that monotonicity is violated by many natural voting methods as is illustrated below.

To get a more complete impression of the implications of monotonicity, let us now explore in some detail whether it is satisfied by some specific well-known social choice correspondences.

Let us begin with the classical exchange environment referred to in Hurwicz (1972) and consider whether the Walrasian correspondence satisfies monotonicity. In Example 1 monotonicity was satisfied. However, that was a very restrictive setting and in fact monotonicity is not satisfied by the Walrasian correspondence on the full classic domain of preferences. This was shown in the context of the following example by Hurwicz et al. (1995).

Example 2 (Hurwicz et al. 1995). Consider a two person, two good Edgeworth economy.[21] There are two states (R^1, R^2) and $(\bar{R}^1, R^2)$, with the preferences as pictured in Fig. 4 above. The allocation a is a Walrasian equilibrium at R^1, R^2, but not at $\bar{R}^1, R^2$.

The difficulties that appear in the figure above have to do with the boundary of the Edgeworth box. In order to satisfy monotonicity, we need to find b such

[21] Although the figure is simply for 2 individuals, it is easily adapted for 3 or more individuals.

that aR^1b and $b\bar{P}^1a$. The allocation b necessary to satisfy this requires an amount of good 2 for player 1 that exceeds the total amount of good 2 available in the economy. The problem is that the crossing of preferences required by monotonicity takes place outside of the Edgeworth box, and outside of the set of allocations that are feasible for a mechanism to provide.

The way around this problem suggested by Hurwicz et al. (1995) is to modify the Walrasian correspondence to include a as an equilibrium at $\bar{R}^1, R^2$. This new correspondence, called the "constrained Walrasian correspondence" satisfies monotonicity and is Nash implementable when there are 3 or more individuals.

To be precise let $A = \{x \in R_+^{N\ell} \mid \sum_i x_k^i \leq \sum_i e_k^i \ \forall k\}$. Say that x is a *constrained Walrasian equilibrium allocation* if there exists $p \in \mathbb{R}_+^{\ell}$ such that $p \cdot x^i \leq p \cdot e^i$ for each i and xR_iy for all $y \in A$ such that $p \cdot y^i \leq p \cdot e^i$.

The constraint in the definition is in the requirement that $y \in A$. In the usual definition of a demand correspondence, one requires that a demanded point be weakly preferred to all other points in the budget set. In this constrained definition, it is only required that a demanded point be weakly preferred to the points in the budget set that are also feasible given the resources in the economy. Thus, a^1 is not in the demand correspondence of player 1 at $\bar{R}^1$, but it is in the constrained demand correspondence. So, the allocation a is a constrained Walrasian equilibrium at $\bar{R}^1, R^2$, and so the difficulty with monotonicity is overcome. Hurwicz, Maskin, and Postlewaite (1995) show that the constrained Walrasian correspondence is monotonic on a classical domain of preferences, and it is contained in any upper-hemicontinuous social choice correspondence that is Pareto efficient, individually rational, and monotonic.

Next, let us examine monotonicity in a voting context.

Example 3. Let A be a finite set of candidates, and consider the following preference profile, with alternatives listed in order of preference.

$R^1 = a, b, c$, $R^2 = c, a, b$, and $R^3 = b, c, a$.

Also consider $\bar{R}^3 = c, b, a$.

Let $F(R) = \{a, b, c\}$ and $F(R^{-3}, \bar{R}^3) = \{c\}$. F agrees exactly with the choice of most scoring rules (e.g., Borda), and most any rule that is anonymous and neutral (e.g., the top cycle correspondence, uncovered correspondence, Copeland rule, Simpson rule, etc.), as well as plurality rule. Unfortunately, F is not monotonic since the standing of a has not changed at all in anyone's preference ranking and yet a is in $F(R)$, but not in $F(R^{-3}, \bar{R}^3)$.

This means that none of the voting correspondences that are commonly used in political elections are Nash implementable![22] This suggests that sequential rationality or other solution concepts are needed for the implementation of voting correspondences.

To see the restrictiveness of monotonicity in the voting context from

[22] Some well-known (but often quite "thick") correspondences such as the Pareto correspondence or union of favorite alternatives, however, are monotonic and Nash implementable.

another perspective, note that the Muller-Satterthwaite (1977) theorem states that if A is a finite set, $\mathscr{P}$ includes all strict preference orderings on A, and $N \geq 3$ then any social choice function that is monotonic is necessarily dictatorial (where a dictator is a single agent who picks his favorite alternative from the range of the function).

As we have seen, although monotonicity can be satisfied, it is in general a strong condition. Many interesting and desirable social choice correspondences are not Nash implementable. In view of this, various alternatives to Nash implementation have been investigated in detail. One possibility is to use alternative equilibrium concepts, an idea we will turn to later. Generally, using stronger equilibrium concepts (i.e., more refined ones) allows for more things to be implemented. Usually the more difficult challenge in implementation comes from ruling out undesired equilibrium outcomes, rather than ensuring that desired outcomes are equilibrium outcomes. Stronger solution concepts aid in ruling out undesired equilibria. The second possibility that has been explored is to place additional structure on the set of alternatives and then to require some sort of approximate, or what has become known as *virtual* implementation. I discuss this next.

4 Virtual implementation

Virtual implementation works under the assumption that if one cannot exactly implement a desired correspondence, then one should be willing to implement a correspondence that is arbitrarily close to the desired one. In order to define what is meant by "close", we need additional structure on the set of alternatives. The way in which this has been done in the virtual implementation literature is to let A be the set lotteries over some primitive set of alternatives. Then we can say that we are implementing something "close" to a desired alternative if we end up with a lottery that places sufficiently high probability on the desired alternative.

If we cannot Nash implement a correspondence we can ask whether there is a correspondence sufficiently close to it, in the lottery sense, that can be Nash implemented. More pragmatically, if a correspondence is not monotonic, is there a correspondence close to it that is monotonic? There is no reason to expect that the answer to this question should be yes. However, the work on virtual implementation has coupled this approximation question together with additional conditions. The answer turns out to be yes if one restricts attention to certain types of correspondences called "ordinal" ones, or if one restricts attention to situations where individuals' preferences have von Neumann-Morgenstern representations. Both of these additional conditions place significant structure on the problem as we shall see and discuss below.

Following with the literature on virtual implementation, let us assume in this section that there is some underlying set of alternatives A' that is finite, and that the set alternatives A is the set of lotteries on A'. More formally, let K denote the number of alternatives in A'. Then $A = \{x \in [0,1]^K \ s.t. \ \sum_k x_k = 1\}$.

This structure on the set of alternatives allows us to define distance among two elements of A through Euclidean distance. This induces a notion of closeness of social choice correspondences:

Two social choice correspondences F and H are ε-close if for every $R \in \mathscr{P}$ there exists a bijection (one to one and onto function) $\tau : F(R) \to H(R)$ such that $|a - \tau(a)| < \varepsilon$ for each $a \in F(R)$.

So, two correspondences are ε-close if for each preference profile there is a one-to-one mapping so that each lottery in one correspondence is within ε of its corresponding lottery in the other correspondence. While this may seem like an innocuous definition, it may be that two lotteries that are ε close to each other are not even roughly equivalent to a society that cares about ex-post realizations. This is a discussion that we will return to after seeing how this definition is used.

A social choice correspondence F is *virtually implementable* if there exists a social choice correspondence G that is ε-close to F and is Nash implementable.[23]

We can now define the ordinality condition mentioned above, which is an invariance condition on F.

Any given preference relation, R_i on A, allows us to compare pure outcomes, which correspond to lotteries that provide probability 1 to some alternative, thereby inducing a preference relation on the set of underlying alternatives A'. For any R^i, denote this preference relation by R^i_p. In what follows, assume that R^i_p is a strict ordering.[24]

F is *ordinal* if $F(R) = F(\bar{R})$ whenever $R^i_p = \bar{R}^i_p$ for all i.[25]

A preference relation R^i is *monotone* if there exists an ordering of A', $a_1, \ldots, a_K$ under R^i_p such that

(i) $a_1 P^i_p a_2 \ldots P^i_p a_K$, and
(ii) if $x \in A$ and $y \in A$ are such that $\sum_{k \le M} x_k \ge \sum_{k \le M} y_k$ for any $M \le K$ and $x \ne y$ then $xP^i y$.

Thus, R^i is monotone (not to be confused with monotonic) if a lottery that puts more weight on more preferred alternatives is preferred to one that puts comparatively less weight on more preferred alternatives. This is a condition that is satisfied by preferences that have a von Neumann-Morgenstern representation, but is more general and allows for a richer set of preferences.

Now we are ready to state a result that contains the main insight of the virtual implementation literature.

---

[23] One can define other notions of virtual implementation by substituting some other solution concept in the place of Nash equilibrium.

[24] This simplifies the exposition. One can weaken this somewhat, as should be clear in what follows, although one cannot allow for complete indifference. Details can be found in Abreu and Sen (1991).

[25] Ordinal is a bit of a misnomer. It requires that F ignore information contained in how preferences vary on A rather than just on A'. Such information is not necessarily "cardinal" information.

Theorem 3 (Matsushima 1988, Abreu and Sen 1991).

- *Any ordinal social choice correspondence defined on a domain of monotone preferences is ε-close to a social choice correspondence that satisfies monotonicity.*
- *Any social choice correspondence defined on a domain of von Neumann-Morgenstern preferences*[26] *is ε-close to a social choice correspondence that satisfies monotonicity.*

Let me outline a proof for each of these results, as the ideas are fairly straightforward. A formal proof, although notationally heavy, can be easily fleshed out by the reader.

First, let $\hat{c}$ be the lottery that places equal weight $\frac{1}{K}$ on every alternative in A'. We construct G as follows. For any $a \in A$ consider the lottery $(1-\varepsilon)a + \varepsilon\hat{c}$. Transforming any lottery to this corresponding lottery we are sure to have an "interior" lottery, or one that puts weight on each alternative. This mapping defines a G that is ε-close to F. The proof is completed by showing that given either of the restrictions, such an interior G[27] is necessarily monotonic. Consider any R, and $\bar{R}$, and $(1-\varepsilon)a + \varepsilon\hat{c} \in G(R)$ such that $(1-\varepsilon)a + \varepsilon\hat{c} \notin G(\bar{R})$. First let us treat the case where F is ordinal and preferences are monotone. Since $G(R) \neq G(\bar{R})$, it follows from the ordinality of F that we can find i such that $R_p^i \neq \bar{R}_p^i$. So there are alternatives $d \in A'$ and $\bar{d} \in A'$ such that $dP_p^i\bar{d}$ and $\bar{d}\overline{P}_p^i d$. From the definition of $\hat{c}$, the lottery $b = (1-\varepsilon)a + \varepsilon\hat{c} + \frac{1}{2K}\bar{d} - \frac{1}{2K}d$ is well-defined and in fact interior.[28] Since preferences are monotone, it follows that $(1-\varepsilon)a + \varepsilon\hat{c}P^i b$ and $b\bar{P}^i(1-\varepsilon) + \varepsilon\hat{c}$. Thus, G is monotonic.

For the case where preferences satisfy the von Neumann-Morgenstern axioms, the proof is quite similar. Note that if $G(R) \neq G(\bar{R})$, then there is some i for whom $R_i \neq \bar{R}_i$. Then, given the linearity of von Neumann-Morgenstern preferences, we can find lotteries $e \in A$ and $\bar{e} \in A$ such that $eP_i\bar{e}$ and $\bar{e}\bar{P}_i e$. This is pictured in the following figure for the case where A' consists of three alternatives.

Letting $b = (1-\varepsilon)a + \varepsilon\hat{c} + \frac{1}{2K}\bar{e} - \frac{1}{2K}e$ again fulfills the requirements to ensure that G is monotonic.

In this setting, given that every social choice correspondence satisfying suitable conditions is close to one that is monotonic, is any such social choice

[26] Here the social choice correspondence is defined on the domain of preferences and not on the domain of utility functions. If one wants to use the set of von Neumann-Morgenstern utility functions as the domain, then the social choice correspondence must be constant across utility functions that are affine transformations of each other.

[27] The reader can check that the following argument is easily adaptable to show that any interior social choice correspondence is monotonic on these domains.

[28] Here d and $\bar{d}$ represent the lotteries given weight one to d and $\bar{d}$. These convex combinations can be made directly as A is a Euclidean simplex.

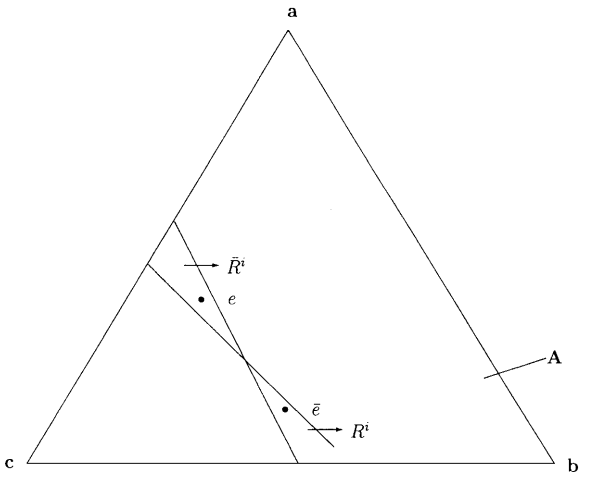

Fig. 5.

correspondence virtually implementable? The answer is yes when $N \geq 3$.[29] One can easily modify the environment so that no veto power is always satisfied. Simply take A to be the set of all lotteries on A', but omitting each pure lottery (that places weight one on some alternative). Then no veto power is satisfied trivially, as there is no best lottery for any individual since with monotone preferences their favorite lottery is a pure lottery. The above theorem then leads to the following corollary.

Corollary 1 (Matsushima 1988; Abreu and Sen 1991). *Let $N \geq 3$. Any ordinal social choice correspondence defined on a domain of monotone preferences is virtually implementable. Any social choice correspondence defined on a domain of von Neumann-Morgenstern monotone preferences is virtually implementable.*

While virtual implementation provides for a remarkable conclusion, it comes at the expense of strong assumptions.

First, in considering virtual implementation, one is implicitly assuming that society is willing to settle for implementing something that is ε-close to the desired social choice correspondence. This raises the question of the ex-post rationality of virtual implementation.[30] Virtual implementation requires that we play with small probabilities on lotteries that may have nothing to do with the desired social choice correspondence, but can be used to distinguish preferences (the d's and e's in the above proof). In particular, with small

[29] For the case of $N = 2$ one needs to add a non-empty lower intersection condition which corresponds to the extra condition needed for two-person Nash implementation. Details appear in Abreu and Sen (1991).

[30] Many mechanisms for implementation can be criticized for the same reason. This applies to the mechanisms for subgame perfect and undominated Nash implementation presented in Sect. 6, and is something I discuss in more detail in Sect. 7.

probabilities these may turn out to be the outcome of the mechanism. In order for the virtual implementation arguments to be valid, agents must take these small probabilities seriously and base decisions on them, with full expectations that these outcomes will be enforced if they happen to be selected by the lottery. It is perfectly possible that an outcome which is bad for all agents is randomly selected, and the arguments underlying virtual implementation require that the agents cannot change this outcome (for example, by renegotiating to some other outcome).

Second, the assumptions of von Neumann-Morgenstern preferences or of ordinality coupled with monotone preferences are strong ones and critical to the arguments. Von Neumann-Morgenstern preferences are assumed in many economics models because of their nice linearity with respect to probabilities, which allows for a very tractable analysis. However, we should always be cautious in interpreting results that rely *critically* on that linearity, rather than just using it for tractability. In other words, if our results are robust to variations in the set of preferences then working with von Neumann-Morgenstern preferences for tractability's sake is fine. However, if the results depend in a special way on a restriction to the domain of von Neumann-Morgenstern preferences, then we should be cautious in applying the results. Arguably, virtual implementation relies critically on the linearity of preferences (or ordinality of F coupled with monotone preferences[31]) because it implies that the crossing conditions needed for monotonicity will be satisfied at all interior points! Once one allows for slight amounts of non-linearity in preferences over lotteries, then the crossing conditions are not automatically satisfied, and monotonicity once again becomes restrictive.

5 Refinements

Another avenue in the theory is implementation under various refinements of Nash equilibrium.[32] As we saw in the Nash implementation theorems, it is easy to ensure that a given desired outcome is an equilibrium outcome while it is difficult to rule out undesired equilibria. This is where the necessity of the monotonicity condition arises and what limits our ability to implement.

[31] Ordinality coupled with the monotone preference condition plays a similar role to the linearity. If $F(R) \neq F(\bar{R})$, then some agent's preferences must differ over at least two underlying alternatives in A'. Then at any interior lottery there must be a crossing of this agent's indifference curves, as substituting a slight amount of one of these two underlying alternatives for the other will be ranked in opposite ways by the two preferences.

[32] When I use the word refinements, I use it literally to mean a solution concept which always selects a subset of Nash equilibria, such as undominated Nash equilibrium, subgame perfect equilibrium, coalition-proof Nash equilibrium, strong equilibrium, and trembling hand perfect equilibrium. Implementation in other solution concepts that are not refinements of Nash equilibrium, such as protective criterion or iterative elimination of weakly dominated strategies also appear in the literature (e.g., Barbera and Dutta 1982 and Abreu and Matsushima 1994), but are not discussed here.

Looking at refinements helps in ruling out undesired equilibria and leads to more permissive results. However, with such additional power will come some cautions about how the refinements are exploited.

Let us begin with a discussion of implementation in undominated Nash equilibrium, as was first explored by Palfrey and Srivastava (1991). This is a simple and natural refinement of Nash equilibrium, that adds to Nash equilibrium the requirement that no individual play a weakly dominated strategy.[33] This solution concept is also useful for illustrating the power of implementation via a refinement of Nash implementation, and some of the cautions that go along with it.

Given a mechanism (M, g), an individual i, and a preference relation for i, R_i, a message $m_i \in M_i$ is *weakly dominated* by $\bar{m}_i \in M_i$ at R_i if $g(\bar{m}_i, m_{-i}) R_i g(m_i, m_{-i})$ for all $m_{-i} \in M_{-i}$ and $g(\bar{m}_i, m_{-i}) P_i g(m_i, m_{-i})$ for some $m_{-i} \in M_{-i}$. A message $m_i \in M_i$ is *undominated* at R_i if there is no other message which weakly dominates it. Let

$$UD(M, g, R) = \{m \mid m_i \text{ is undominated at } R_i \ \forall i\}.$$

A profile of messages $m \in M$ is an *undominated Nash equilibrium* of (M, g) at $R \in \mathscr{P}$ if $m \in NE(M, g, R) \cap UD(M, g, R)$. So let $UNE(M, g, R) = NE(M, g, R) \cap UD(M, g, R)$.

The definition of undominated Nash implementation is then that $F(R) = O_{UNE}(M, g, R)$ for all $R \in \mathscr{P}$.

In order to keep the discussion of undominated Nash implementation relatively uncluttered, let us impose the following assumption on the domain of preferences. A slightly weaker value distinction condition appears in Palfrey and Srivastava (1991), and the reader is referred there for details. This condition allows us to provide a unified exposition of a number of results.

Strict value distinction. The domain $\mathscr{P}$ satisfies strict value distinction if for every $R \in \mathscr{P}$ and $\bar{R} \in \mathscr{P}$ with $\bar{R} \neq R$:

(I) For all i, there exists $a \in A$ and $b \in A$ such that aP_ib,
(II) For all i, if $R \neq \bar{R}_i$ there exists $a \in A$ and $b \in A$ such that aP_ib and $b\bar{P}_ia$.

(I) simply rules out complete indifference, and (II) says that if an individual's preferences change, then they change in a non-trivial way so that *some* alternatives switch ranking (rather than just becoming indifferent). So (II) simply requires that if preferences change then indifference curves cross somewhere, which is a condition that is vacuously satisfied in most contexts of interest. For instance, strict value distinction is satisfied on a domain of strict preferences (linear orders), as well as any setting where preferences are

[33] Such equilibria are a superset of trembling hand perfect equilibria (except for $N = 2$ where they coincide, see van Damme 1987), and thus have nice existence properties, but also have the advantage of being easier to define and work with in such abstract environments as those arising in the implementation literature.

upper-semi continuous and locally non-satiated[34] and so applies to almost any setting of interest.

Theorem 4 (Palfrey and Srivastava 1991). *If $N \geq 3$ and strict value distinction is satisfied, then any social choice correspondence that satisfies no veto power is implementable in undominated Nash equilibrium.*

The contrast of this theorem with Theorem 2 is dramatic, as any social choice correspondence in economic settings as well as most voting rules are admitted. The refinement to undominated Nash equilibrium has thus made a big difference in which social choice correspondence are implementable, completely eliminating the necessity of monotonicity. The main intuition is that any switch in preferences (one is guaranteed to exist by value distinction) can be used to change the set of strategies that are undominated from one state to another. The central difficulty in Nash implementation that is handled by the monotonicity condition is that there can be an alternative a that one wants to be a Nash equilibrium outcome in one state but not in another state. So one needs a preference reversal between a and some other alternative b that can be obtained through a deviation. When one considers undominated Nash implementation, one instead needs a to be supported as an undominated Nash equilibrium outcome in one state R but not in another $\bar{R}$. Let m be an undominated Nash equilibrium message profile at R of some mechanism. In order to rule m out as an undominated Nash equilibrium at $\bar{R}$, it must be that either m is no longer a Nash equilibrium or some m^i is weakly dominated at $\bar{R}^i$ (while it was undominated at R^i). One can assure that m^i is weakly dominated at $\bar{R}^i$ and not at R^i, simply by taking advantage of any difference in the preferences between R^i and $\bar{R}^i$. Of course, one has to be careful in working out the details so that such differences are accounted for in each state while still producing the desired undominated Nash equilibria in each state, which results in a complex mechanism.

The idea that domination arguments can take advantage of relatively slight differences in preferences,[35] can be pushed further. In fact an even more permissive theorem holds with a weaker solution concept. We can drop no veto power, weaken the solution concept to undominated strategies, and include the case of $N = 2$.

Theorem 5 (Jackson 1992). *If strict value distinction is satisfied, then any social choice correspondence is implementable in undominated strategies.*

The proof of each of these theorems involves mechanisms that are quite intricate. The underlying principles are similar to that of the mechanism used to prove Theorem 2: if there is agreement in the announcements then the out-

[34] See Jackson (1992) for a proof.

[35] You may notice some analogies to way in which slight differences in preferences over lotteries are expoited by virtual implementation. The same criticisms regarding the ex-post rationality of enforcing the outcomes can be made, and I discuss this in more detail below.

come is implemented (as in (1) in the mechanism proving Theorem 2), there is a possibility for some individuals to deviate and allow for some "test" pair of alternatives which serve the role of changing which strategies are weakly dominated, and there is some use of an integer game (as in (3)) to rule out certain undesired configurations of strategies. However, the mechanisms are more complex in terms of how and when the test pairs apply in order to take advantage of weak dominance arguments. Rather than exhaust the reader with the details of the mechanisms, I present a simple example that illustrates an implementing mechanism for Theorem 5 in a specific case. The example is also useful for another purpose: it points out a serious caveat that we should have regarding implementing mechanisms.

Example 4 (Jackson 1992). Let $N = 2$ and $A = \{a, b\}$. $R_1 = R_2$ is such that aP_ib, and $\bar{R}_1$ is such that $b\bar{P}_1a$.

Consider $F(R_1, R_2) = \{b\}$ and $F(\bar{R}_1, R_2) = \{a\}$. This is a peculiar social choice correspondence as it goes exactly against the preferences of the agents. But it is then convincing[36] that if we can implement this social choice correspondence then we can implement any social choice correspondence. The following mechanism does the trick.

	m^2	$\tilde{m}^2$	$\hat{m}^2$									
m^1	b	a	a	a	a	$\cdots$	a	a	a	a	a	$\cdots$
	b	a	a	a	a	$\cdots$	b	b	b	b	b	$\cdots$
	b	b	a	a	a	$\cdots$	b	b	b	b	b	$\cdots$
	b	b	b	a	a	$\cdots$	b	b	b	b	b	$\cdots$
	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$		$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	
$\bar{m}^1$	a	b	b	b	b	$\ldots$	b	b	b	b	b	$\ldots$
	a	a	a	a	a	$\ldots$	a	b	b	b	b	$\ldots$
	a	a	a	a	a	$\ldots$	a	a	b	b	b	$\ldots$
	a	a	a	a	a	$\ldots$	a	a	a	b	b	$\ldots$
	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$		$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	

In this table, the $\cdots$ and $\vdots$ indicate a countable continuation of the series of entries and corresponding strategies.[37] Let us check that the mechanism implements the stated social choice correspondence. First, m^2 is undominated for 2 since it is the only strategy that results in a against $\bar{m}^1$. Second, any strategy in the right half of the mechanism for individual 2 is dominated by individual 2's strategy $\tilde{m}^2$. Next, $\tilde{m}^2$ is dominated by $\hat{m}^2$. Next, $\hat{m}^2$ is dominated by the strategy immediately to its right; and so forth. This leaves m^2 as

[36] The hard-to-convince reader is referred to Jackson (1992) for the proof of Theorem 5.
[37] Each individual here has a message set that is divided into two sets that are both countably infinite. So for instance, m^2 gets b against every message in 1's first set and a against every message in 1's second set.

the only undominated strategy for individual 2. A similar analysis for player 1 leads to a unique undominated strategy of m^1 at R^1, and of $\bar{m}^1$ at $\bar{R}^1$. Thus, the mechanism implements the stated F.

The elimination of dominated strategies is a questionable practice in the above mechanism. For instance, in the sequence $\tilde{m}^2, \hat{m}^2, \ldots$, each strategy is eliminated by the next, but there is no undominated or 'best' strategy in that string. We end up predicting that individual 2 will not play any of these strategies, but will play m^2 instead, even though it does worse in some situations than any of the eliminated strategies.[38]

As shown in Jackson (1992) this is not simply a problem when dealing with implementation in undominated strategies, but also holds for undominated Nash implementation. This example can be modified to have three individuals and implement essentially the same social choice function in undominated Nash equilibrium, via a mechanism with similar difficulties. Moreover, in each of the examples the same is true of every implementing mechanism.

To address the questionable removal of dominated strategies in Example 4, we can rule out the use of such mechanisms by requiring that the implementing mechanism be *bounded*. This requires that when a strategy is dominated, it be dominated by some undominated strategy.[39]

5.1 Bounded mechanisms

A mechanism (M, g) is *bounded* relative to $\mathscr{P}$ if for every i and $R \in P$, whenever $m_i \in M_i$ is dominated at R_i, then it is dominated by some message $\bar{m}_i \in M_i$ that is undominated at R_i.

This definition makes reference to the preference domain, and so boundedness is a condition relative to an environment. A finite mechanism is bounded regardless of the environment due to the transitivity of the domination relation. However, an infinite mechanism's boundedness property may depend on the domain of preferences in question.

If we require implementation via a bounded mechanism, then Theorem 5 no longer holds and necessary conditions arise.

F is *strategy-resistant* if for all i, $R \in \mathscr{P}$ and $\bar{R}_i$ such that $(\bar{R}_i, R_{-i}) \in \mathscr{P}$ and for each $b \in F(\bar{R}_i, R_{-i})$, there exists $a \in F(R)$ (possibly $a = b$) such that aR_ib.

---

[38] As an aside, the use of the integer game in the mechanism used for Nash implementation (i.e., proving Theorem 2) may have bothered you. As no domination arguments were used there, the issues are slightly different than that in Example 4. However, the integer game used in proving Theorem 2 had a similar feature in that it could produce situations where agents have no best response (i.e., if they have beliefs that have positive probability that some other individual will use any integer). I will return to discuss this issue after discussing the domination issues.

[39] An alternative approach, would be to modify the concept of domination so that weakly dominated strategies are eliminated only if they are dominated by some undominated strategy. (Timothy Van Zandt suggested this alternative approach in a discussion of Jackson 1992). As is easily seen from the proof of Theorem 6, the same restrictions on implementation ensue with either approach.

This condition has a flavor of strategy-proofness, but is defined for correspondences. If we think of i at R, considering a manipulation by acting in accordance with $\bar{R}_i$, we find that for whatever outcome i may hope for, $b \in F(\bar{R}_i, R_{-i})$, there is some outcome $a \in F(R)$ which is at least as good as b. This condition is very strong: it reduces to strategy-proofness when F is a function.

The following theorem illustrates the impact of requiring boundedness in implementation via undominated strategies.

Theorem 6 (Jackson 1992). *If a social choice correspondence is implementable in undominated strategies via a bounded mechanism, then it is strategy-resistant.*

The proof of this theorem is quite easy. Let F be implemented in undominated strategies via the bounded mechanism (M, g). Consider i, $R \in \mathscr{P}$ and $\bar{R}_i$ such that $(\bar{R}_i, R_{-i}) \in \mathscr{P}$ and consider any $b \in F(\bar{R}_i, R_{-i})$. Since F is implemented, there exists $\bar{m} \in UD(M, g, \bar{R}_i, R_{-i})$ such that $g(\bar{m}) = b$. Either $b \in F(R)$, in which case let $a = b$, or it must be that $\bar{m}_i$ is dominated by some $\tilde{m}_i$ that is undominated at R_i. Let $a = g(\tilde{m}_i, \bar{m}_{-i})$. Since $(\tilde{m}_i, \bar{m}_{-i})$ is undominated at R, it follows that $a \in F(R)$, and from the definition of domination, it follows that that aR_ib.

Thus, we end up with a stark contrast between what is implementable in undominated strategies when we can use any mechanism and when we can only use bounded mechanisms. This suggests it is very important to understand the restrictions imposed by considering only mechanisms for which a given solution concept is appropriate.

Boundedness ends up restricting the class of social choice correspondences that are implementable in undominated Nash equilibrium as well. That is studied in Jackson et al. (1994). It turns out to introduce a necessary condition (the chained condition) that is weaker than monotonicity and strategy-resistance, but nonetheless rules out some well-known correspondences. The reader is referred to that paper for details.

While boundedness is a condition that is natural to require when we examine elimination of dominated strategies, or solution concepts like undominated Nash equilibrium or iterative elimination of weakly dominated strategies, it is not an obvious condition to require when considering Nash equilibrium, or subgame perfect equilibrium, where the elimination of dominated strategies is never an issue. However, there are conditions that one might like to impose on mechanisms so that such solution concepts are reasonable.

5.2 Mixed strategies

The definition of Nash equilibrium that we worked with in proving Theorem 2, does not consider mixed strategies. Even giving a proper definition of mixed strategy equilibrium requires additional structure as preferences have to be defined over lotteries on alternatives, and since message spaces may be uncountably infinite one has to take care. Without tackling these technical issues, let us

consider an example that shows that worrying about the possibility of mixed-strategy equilibria makes a difference.

What we can show is that if one Nash implements a social choice correspondence by a finite mechanism, then there may still exist mixed strategy Nash equilibria that result in outcomes outside of the social choice correspondence and which may be preferable from the individuals' perspective.

Example 5 (Jackson 1992). Let $N = 2$ and $A = \{a, b, c, d\}$. Preferences are described by $aP^1bP^1cI^1d$ (where I^1 indicates indifference), $aP^2bP^2cP^2d$ and $b\bar{P}^2a\bar{P}^2c\bar{I}^2d$.

Consider F such that $F(R_1, R_2) = \{a\}$ and $F(R_1, \bar{R}_2) = \{d\}$. F is Nash implemented by the following mechanism.

	m^2	$\hat{m}^2$	$\tilde{m}^2$
m^1	d	c	c
$\hat{m}^1$	d	a	b
$\tilde{m}^1$	d	b	a

Note however, that at $(R^1, \bar{R}^2)$ for most definitions of preferences over lotteries, there also exists a mixed strategy equilibrium that results in a and b with positive probability. This is true for every finite implementing mechanism.

Claim 1. *Suppose that players' preferences over lotteries on A are continuous, convex, and monotone. Any mechanism with a finite number of strategies for each player has a Nash equilibrium in mixed strategies at $R^1, \bar{R}^2$ resulting in outcomes a or b with positive probability, and which strictly Pareto dominates the outcome d.*

The proof of the claim is simple. Suppose that a mechanism (M, g) Nash implements F. Consider $\hat{M}^i \subset M^i$ such that $\hat{m}^i \in \hat{M}^i$ implies that there exists $\hat{m}^{-i}$ such that $g(\hat{m}) \in \{a, b\}$. Thus, $\hat{M}^i$ is found by eliminating strategies that can only ever lead to c or d. Since F is implemented, there must be some strategy profile $\hat{m}$ such that $g(\hat{m}) = a$, and so $\hat{M}$ is non-empty. Given the assumptions on preferences and the finiteness of the mechanism, there exists a Nash equilibrium in mixed strategies to $(\hat{M}, g)$ at $R^1, \bar{R}^2$. Given the definition of $\hat{M}$ and the fact that preferences on lotteries are monotone, it follows that this equilibrium must put positive probability on at least one of the outcomes a and b, as these are reachable by some strategy of 1 given any (mixed) strategy of 2, and are preferred by 1 to c and d over which 1 is indifferent. Note that $\hat{m}$ is then also a Nash equilibrium of (M, g) at $R^1, \bar{R}^2$ since strategies in $M^i \backslash \hat{M}^i$ only lead to c or d and cannot provide improving deviations. This establishes the claim.

The above claim shows that the standard definition of Nash implementation that only considers pure strategies can be a problematic one. Let me make two remarks about the above example.

First, the social choice correspondence that is being implemented is a bit strange in that we are trying to implement d when it is a least preferred outcome for both of the individuals. However, this might be quite natural in the

context of a problem such as a principal/multiple-agent problem, where the two individuals above represent the agents, the principal designs the mechanism and prefers that the agents take actions leading to outcome a in the first state and d in the second state.

Second, the finiteness of the mechanism is critical to the result. If we allowed for infinite mechanisms, then we could ensure that there are no mixed strategy equilibria, by employing an integer game.[40] Maskin (1999) shows that one can design a mechanism for Nash implementation that has no mixed strategy equilibria (see the Appendix of his article); but his mechanism involves an integer game and thus an infinite message space. Example 5 shows that there are correspondences in finite settings that are Nash implementable when one only considers pure strategies, but not Nash implementable (except by infinite mechanisms) when one allows for mixed strategies.

For Example 5 it is natural to require finite mechanisms as the setting is completely finite. However, many settings of interest are naturally infinite and so requiring implementation in finite mechanisms might not always make sense. So one might ask, "What is the correct restriction on mechanisms is for Nash implementation if one accounts for mixed strategies?" It should be a condition that naturally rules out improper use of integer games, without unduly restricting the message space or outcome function when infinite spaces might be appropriate.

The bothersome aspect about integer games is that they are appended to mechanisms precisely because there is no Nash equilibrium to them, even though players could very well try to announce higher integers than others in the hopes that they will get their most-preferred alternative. We have no prediction for how a player will act if he or she believed that the other players would be announcing integers. A player's best response correspondence is not well-defined when that player faces a mixed strategy of the others that places weight on an infinite set of integers. So, one way to rule this out is to look only at mechanisms for which best response correspondences are well-defined. Jackson et al. (1994) suggest such a requirement, calling it the "best response property." The best response property is satisfied by any finite mechanism and also by many infinite mechanisms. Nevertheless, it is a strong condition and rules out some mechanisms that are well-behaved in the way that they either admit or rule out mixed strategy equilibria.

The same point regarding mixed strategies in this example that was made in Claim 1 above, can be made with regard to subgame perfect implementation.[41]

---

[40] We saw such a construction in the mechanism used to prove Theorem 2. Under (3), the outcome was that announced by the agent who named the highest integer. Clearly as long as there is some disagreement over most preferred outcomes, then there is no equilibrium in (pure or) mixed strategies to the integer game.

[41] Moore and Repullo (1988) provide a mechanism for subgame perfect implementation that has no mixed strategy equilibria. However their mechanism also employs an integer game and fails the best response property.

Claim 2. *Suppose that players' preferences over lotteries on A are continuous, convex, and monotone. Any extensive form mechanism*[42] *that has a finite number of terminal nodes and that implements F, has a subgame perfect equilibrium in mixed strategies at* $R^1, \bar{R}^2$ *which results in outcomes a or b with positive probability and which strictly Pareto dominates the outcome d.*

The proof of Claim 2 is similar to that of Claim 1. We first need to trim the tree to obtain a game form like $(\hat{M}, g)$ in the previous proof. In the extensive form this is a bit trickier, but the finiteness allows us to do this inductively. Start with information sets that only precede terminal nodes, and eliminate actions that only lead to c or d at those information sets. If all actions are eliminated, then replace the information set with terminal nodes that lead to d. Next, proceed to information sets that lead to either terminal nodes or information sets considered in the first step, and perform the same trimming. Proceed inductively until the root is reached. Find a subgame perfect equilibrium of the remaining extensive form. It must lead to a or b with positive probability. Then construct a subgame perfect equilibrium of the original game form leading to the same outcomes, by picking any action at information sets that were completely eliminated in the trimming procedure. This then establishes the claim.

The difficulties with mixed strategies pointed out in the claims above, can be attacked from different directions. One approach to handling this problem is to examine virtual implementation to see whether the difficulties can be avoided there. Abreu and Matsushima (1992a) have analyzed virtual implementation via the iterative elimination of strictly dominated strategies. Since the set of Nash equilibria (pure or mixed) has support in the set of strategies that survives the iterative elimination of strictly dominated strategies, if one obtains a unique profile of actions via the iterative elimination of strictly dominated strategies, then there must be a unique Nash equilibrium (even considering mixed strategies). Abreu and Matsushima show that with three or more individuals, any social choice function is virtually implementable in iteratively undominated strategies by a finite mechanism. Their construction of an implementing mechanism is very clever and quite intuitive. I refer the reader to their paper for details. While they obtain a powerful result handling mixed strategies, it comes at the expense of the strength of the assumptions in the virtual implementation setting (as discussed above).[43]

6 Specific environments

Another approach to accounting for mixed strategies, and also to providing implementation by mechanisms that are bounded/and or simple and natural

[42] The definition of extensive form mechanism is the same as that of an extensive form game, except that there is an outcome associated with each terminal node rather than a payoff. As the definition is standard, available in most any game theory text, and notationally cumbersome, I omit it here.

[43] Glazer and Rosenthal (1992) criticize the Abreu and Matsushima mechanisms on complexity grounds and Abreu and Matsushima (1992c) offer a reply.

in other ways, is to examine implementation in specific environments. Part of the reason that the mechanisms used in the sufficiency theorems are so abstract and complicated is that they implement a given social choice correspondence without using any detailed information about the specific restrictions on preferences that may be satisfied in an environment. If we take advantage of knowledge about the structure of the environment, the implementation problem can be simplified.

One important class of environments to be considered is that of economic environments where there is a private good. In such environments there is a disagreement between agents' preferences that allows a mechanism to play agents against each other – each agent prefers to have more of the private good. Also, there is some possibility of imposing "fines" or punishments that are agent specific. Let us examine two results on implementation in such settings that employ simple and intuitive mechanisms that are bounded and handle mixed strategies. The first such result applies to undominated Nash implementation and environments that are called separable.

6.1 Separable environments

(1) there exists $w \in A$ such that $aP^i w$ for all i, $R \in \mathscr{P}$, and $a \in F(R)$,
(2) for any $a \in A$, $J \subset N$, and $R \in \mathscr{P}$, there exists a^J such that $a^J I^i w$ for any $i \in J$ and $a^J I^i a$ for $i \notin J$,
(3) $R^i \neq \bar{R}^i$ implies that there exist a and b such that $aP^i bP^i w$ and $b\bar{P}^i a\bar{P}^i w$.

(1) is a requirement that there is an outcome that each agent considers to be worse than any alternative in the range of the social choice correspondence. (This means that the definition of a separable environment is relative to a given social choice correspondence.) For instance, in an exchange economy this would be satisfied by setting w to be the 0 allocation, provided the social choice correspondence leaves each agent with some minimal consumption in every state. (2) is the requirement that captures the separable nature of the environment. It states that it is possible to reduce the allocations of some set of individuals J while not affecting the remaining individuals. This is satisfied in exchange economies that satisfy free disposal, as well as other settings such as those with public goods and/or externalities provided there is at least one disposable private good that agents value sufficiently. (3) is a version of strict value distinction again, with the additional requirement that the test pair of alternatives (a and b) be preferred to the bad outcome. Again, it is satisfied in many economic settings.

Theorem 7 (Jackson et al. 1994; Sjöström 1994[44]). *If the environment is separable relative to a social choice function F, then F is implementable in undominated Nash equilibrium by a bounded mechanism that has no mixed strategy equilibria.*

[44] The definition of separable environments is from Jackson et al. (1994). Although Sjöström's (1994) setting is slightly different, the intuition behind his mechanism is identical.

Theorem 7 is proven using the following mechanism.

Each individual i announces either a pair of preference relations $m^i = (R^i, R^{i+1})$ (interpreted as being an announcement i and $i+1$'s preferences respectively, letting $N+1=1$), or a pair of alternatives $m^i = (a, b)$. There are three cases to consider.

(I) All individuals announce pairs of preferences. In this case, let J be the set of i such that $m_2^i = m_1^{i+1}$. This is the set of i whose announcement about $i+1$ matched what $i+1$ announced about him or herself. The outcome is then $[F(m_2)]^{N\setminus J}$.[45]

(II) Some i announces $m^i = (R^i, R^{i+1})$ and all other $j \neq i$ announce $m^j = (a, b)$ such that aR^iw and bR^iw (with all $j \neq i$ announcing the same (a, b)). Then $g(m) = a^{N\setminus\{i\}}$ if aR^ib and $g(m) = b^{N\setminus\{i\}}$ otherwise (where $a^{N\setminus\{i\}}$ and $b^{N\setminus\{i\}}$ are as defined in (2) of separable environments).

(III) For any other configuration of messages $g(m) = w$.

Since an individual's announcement about his or her preferences can only affect his or her allocation in (II), it is clear that the only undominated announcement of preferences is to announce one's own preference truthfully. Also, since by announcing a pair of alternatives one always obtains an alternative equivalent to w, any announcement of a pair of alternatives is dominated by a truthful announcement of preferences. Finally, given that all individuals should be announcing pairs of preferences, and announcing their own truthfully, it then follows from (I) that each individual should announce their neighbor's preference truthfully. This is the unique undominated Nash equilibrium. The mechanism is bounded because the dominated actions were dominated by undominated actions.

Thus, in this separable class of environments we can obtain a strong implementation result with a relatively simple bounded mechanism. Before discussing some of the shortcomings of the mechanism above, let me discuss a closely related setting and mechanism for implementation in subgame perfect equilibrium.

A special case of separable environments is that of quasi-linear preferences. That setting allows for a simple, clean, and intuitive implementation in subgame perfect equilibrium.

Let us say that an environment has *quasi-linear preferences* that are representable by bounded utility functions if there exists A' such that

(1) $A = \{(a, t^1, \ldots, t^N) \in A' \times \mathbb{R}^N \mid \sum_i t^i \leq 0\}$, and

(2) each R^i is represented by a utility function $u^i : A' \to [0, 1]$ such that $(a, t^1, \ldots, t^N)\ R^i\ (b, \bar{t}^1, \ldots, \bar{t}^N)$ if and only if $u^i(a) + t^i \geq u^i(b) + \bar{t}^i$.

Condition (1) says that allocations can be written as a product space of some primitive allocations as well as a vector of private transfers. Note that

[45] So this is the outcome where individuals in J get F evaluated at the R corresponding to that in the second entry of each agent's message (what they announced about their neighbor), and individuals outside of J get an outcome equivalent to the bad outcome w.

the transfers are freely disposable since $\sum_i t^i \leq 0$. Condition (2) states that preferences are representable by a bounded quasi-linear utility function.

Let us consider social choice functions for which $F(R) = (a, 0, \ldots, 0)$ for every $R \in \mathscr{P}$. This is almost without loss of generality, since A' is arbitrary and if one wishes to specify transfers that do not sum to 0, for instance specifying the sharing of the cost of a public good as a function of R, then this can be included in the specification of A' and the t^i's can be thought of as additional transfers. The only loss of generality comes from condition (2) above, which implies that any transfers included in A' must be bounded in size. This bound is needed for the implementing mechanism below to function.

Theorem 8 (Moore and Repullo 1988). *Consider an environment with a finite number of states, quasi-linear preferences represented by utility functions that are bounded, and $N = 2$. Any social choice function is implementable in subgame perfect equilibrium by an extensive game form of perfect information of finite length that has no mixed strategy equilibria.*

Although the theorem is stated for $N = 2$, the mechanism below can be extended to $N > 2$ (see Moore and Repullo 1988 and Moore 1992 for details).

Theorem 8 is established through the mechanism pictured below. To understand the mechanism we note two points. First, the reader can easily check that for any quasi-linear utility functions on A', u^i and v^i that represent different preferences on A, we can find (a, t^i) and $(b, \bar{t}^i)$ with $0 \geq t^i$ and $0 \geq \bar{t}^i$ such that

$$u^i(a) + t^i > u^i(b) + \bar{t}^i \quad \text{while} \quad v^i(a) + t^i < v^i(b) + \bar{t}^i.$$

Thus, in the challenge phase of the mechanism pictured below, if agent i has announced v^i and is challenged by the other agent who counters with u^i, then such a pair is invoked. Second, we can rewrite the social choice function to be a map from profiles of bounded utility functions on A' into allocations. So, the implemented social choice function can be written to depend directly on the announcements of u^1 and u^2.

The intuition behind the mechanism is straightforward. Suppose that the true profile of preferences corresponds to (u^1, u^2). If 1 announces a false preference v^1 instead of u^1, then individual 2 can challenge 1 and say u^1. If 1 is challenged then 1 will end up paying a large fine no matter what occurs in the sequel. 1 is then offered a choice between the (a, t^i) and $(b, \bar{t}^i)$ that distinguish u^1 and v^1. 1 will choose honestly, as this is the end of the tree and 1 pays a fine regardless of the choice. If 1 chooses (a, t^i) then this indicates that 1 was not truthful in the first stage, and so 2 is rewarded by receiving the fine that 1 pays. If 1 chooses $(b, \bar{t}^i)$ then this indicates that 2 was not truthful (either in challenging or in 2's announcement of u^1) and then 2 also pays a fine. This fine is sufficiently large (outweighing any implemented allocation) that 2 would like to challenge 1 if 1 lies, and not otherwise. The same reasoning applies to the other section of the tree and so the unique subgame perfect equilibrium out-

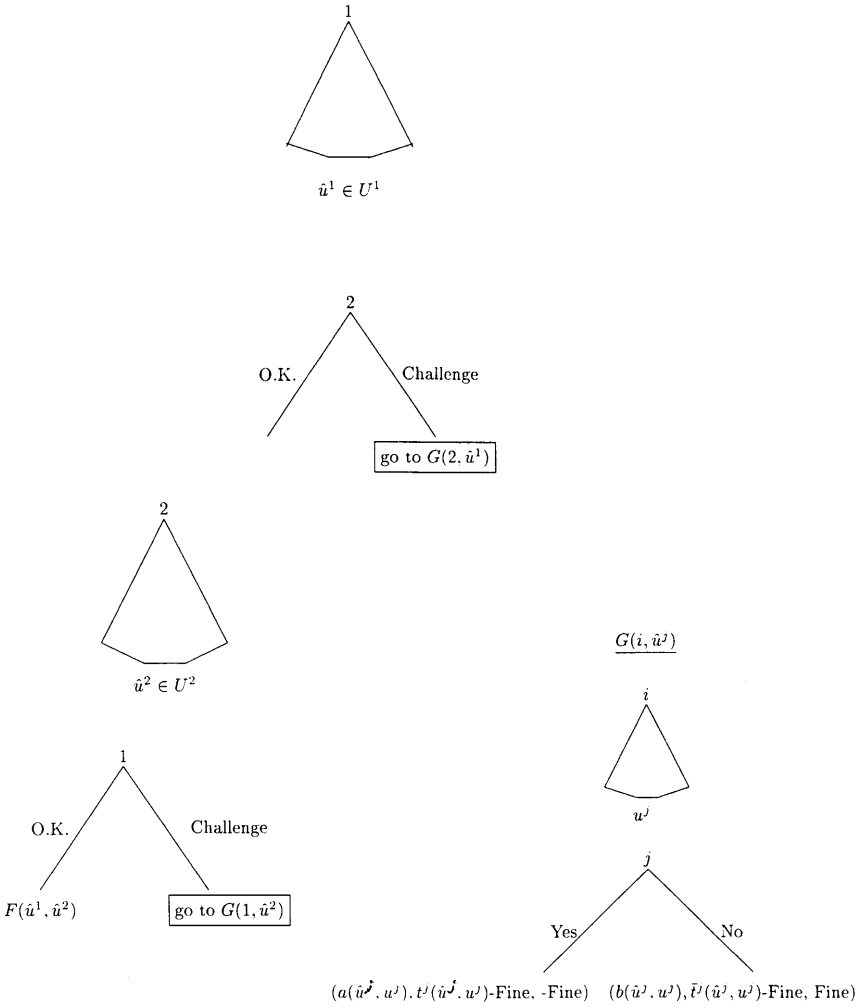

Fig. 6.

come is for both individuals to announce utility functions that represent their true preferences and the correct outcome is implemented.[46]

While the results in this section have very optimistic implications, they come at the expense of strong restrictions. Most importantly, one of the criticisms made of virtual implementation – that it may not be credible to believe that bad outcomes will stand – is also particularly important here. In both of

[46] The finiteness of the state space is important in guaranteeing existence of equilibrium, as it ensures that there exists a well-defined best response of the "challenging" agent in the challenging subgames.

the mechanisms above there is some "bad" outcome (w in the undominated Nash implementation and a fine large enough to outweigh all other considerations in the subgame perfect implementation) and the mechanisms need to commit to these bad outcomes in order to sustain the implementation. There are papers that address this shortcoming, which I discuss below.

Finally, while the mechanisms above are relatively simple and intuitive, bounded, and avoid mixed strategy equilibria; the mechanisms are still cumbersome in the sense that they require announcements of preferences (or utility functions). Thus, although they are well-behaved in a theoretical sense they are still far from "natural" in the sense that they could be easily used in practice. So, although they are useful in delineating the boundary of what can be implemented, we should want to push further to require implementation by a mechanism with a simple or natural message space.[47] By adding even more structure to the problem, one can obtain natural mechanisms with simple action spaces. The tractability of specific environments for obtaining intuitive simple mechanisms for implementation has been taken advantage of in allotment, bargaining, voting, principal-agent, and public goods environments in a series of papers.[48] The implementing mechanisms that are obtained often have auction and voting like features that make them very simple and natural.

One specific environment that has received a great deal of attention in the implementation literature is that of the exchange of private goods, and the implementability of the constrained Walrasian correspondence.[49] Within this context one can examine implementation with respect to certain message spaces. Can the constrained Walrasian correspondence be implemented by mechanisms that only involve announcements of prices and quantities or announcements of prices and allocations? These questions and others are addressed in recent papers by Saijo, Tatamitani and Yamato (1996), Dutta et al. (1995), and Sjöström (1995b), which map out exactly what sorts of message spaces are needed for Nash implementation of the Walrasian correspondence.

---

[47] One approach to addressing the complexity of announcements of the mechanism is to examine the dimensionality of the message spaces and look for mechanisms with minimal sized message spaces and to establish lower bounds on what is necessary for implementation in various environments. For explicit discussion of the size and complexity of mechanisms for decentralization see the seminal paper of Mount and Reiter (1974). Papers focusing on this issue in implementation (when incentives are also an issue) include Reichelstein and Reiter (1988), Saijo (1988), and Hong and Page (1994).

[48] See, for example, Glazer and Ma (1989), Jackson and Moulin (1992), Rubinstein and Wolinsky (1992), Dutta and Sen (1993), Glover (1993), Thomson (1994), Bag (1996, 1997), Jackson and Palfrey (2001), and Brusco and Jackson (1999).

[49] The large literature on this subject begins with Hurwicz (1972), and includes papers that provide increasingly well-behaved mechanisms worrying about auxiliary properties of the mechanism like continuity of the outcome function, individual rationality, and balance out of equilibrium. See for example, Schmeidler (1980), Hurwicz et al. (1995), Wettstein (1992), Nakamura (1990), and Hurwicz (1996).

7 Other topics and open questions

While the discussion of implementation theory above exposes the reader to some of the results, techniques, and themes in the literature, there is a substantial portion of the literature that I have not discussed. In this section I briefly discuss some of the issues that are addressed in the rest of the literature as well as some interesting open problems.

7.1 Implementation under incomplete information

There is an extensive literature on implementation when individuals hold private information. Nash implementation has a natural analog in such settings which is referred to as Bayesian implementation.[50]

When individuals may hold private information (that is not known collectively to the other individuals in the society), an incentive compatibility condition becomes necessary. In a complete information setting if I ask agents to tell me the state and the announcements do not all agree, then I can be sure that someone is lying. In an incomplete information setting, where only one person knows the state, there is no way to check whether that person is lying and so it must be in their interest to announce the state in a direct revelation mechanism. This discussion is in terms of direct revelation mechanisms, but it implies that an incentive compatibility condition is necessary in incomplete information settings that is not necessary (or alternatively is vacuously satisfied) in complete information worlds. As we know from the mechanism design literature, incentive compatibility conditions can turn out to be quite restrictive, rendering many efficient and desirable social choice functions non-implementable.[51]

Once one adds incentive compatibility as a necessary condition, we can use our understanding of implementation in complete information settings to make significant progress on implementation in incomplete information settings. For instance, the reasoning behind the necessity of monotonicity for Nash implementation extends to provide an analogous necessary condition for Bayesian implementation that has been called Bayesian monotonicity in the literature (e.g., see Postlewaite and Schmeidler 1986, Palfrey and Srivastava 1989a, and Jackson 1991[52]). Although the Bayesian monotonicity condition is more complex, the intuition underlying it is similar to that behind monotonicity.

While the necessity of an incentive compatibility and a Bayesian monotonicity condition would be expected[53], the Bayesian setting also introduces

[50] Bayesian implementation may be seen as a generalization of Nash implementation, as it reduces to Nash implementation in situations where agents are symmetrically informed.

[51] See Palfrey and Srivastava (1987) and (1993) for a number of examples in the context of Bayesian implementation.

[52] How one defines the setting and the Bayesian monotonicity condition is important for identifying a condition which is both necessary and sufficient for implementation. See Jackson (1991) for a discussion of this point.

[53] A closure condition which links play of different equilibrium strategies across the common knowledge partition is also necessary, and quite intuitive.

new considerations that render the task of finding sufficient conditions significantly more complex than in the complete information setting. For instance, one cannot state a direct analog of Theorem 2 in a Bayesian environment. In particular, a no-veto style condition and Bayesian monotonicity need to be carefully intertwined respecting the information structure, as discussed in Jackson (1991).[54] There are other interesting differences that arise in a Bayesian setting: Dutta and Sen (1994) demonstrate a simple social choice function in a finite setting that requires a mechanism with infinite message spaces for implementation.

While progress has been made in understanding the conditions that characterize Bayesian implementation, the full implications of the Bayesian monotonicity condition and the extent to which it may be satisfied are still less well understood. Palfrey and Srivastava (1987, 1993) make significant headway in showing that many analyses in mechanism design and agency theory that use incentive compatibility and invoke the revelation principle suffer from multiple equilibrium problems and the identified social choice functions fail to satisfy Bayesian monotonicity condition. They also show that in environments with transferable utility the conditions for Bayesian implementation are more easily satisfied.[55] A recent paper by Serrano and Vohra (1999) sheds more light on the restrictiveness of Bayesian implementation by showing that Bayesian monotonicity is essentially an ordinal condition, and so implemented social choice correspondences must be constant across different cardinal representations of underlying ordinal preferences. What is left open is a detailed understanding of what can be implemented in non-transferable utility settings when we worry not only about incentive compatibility, but also about full implementation and thus the multiple equilibrium problem and Bayesian monotonicity.

While Nash implementation has an obvious generalization to environments with incomplete information,[56] subgame perfect implementation does not have a unique generalization, but instead several. This is due to the variety of alternative formulations of sequential rationality under incomplete information. In particular, varying assumptions about how individuals update beliefs off the equilibrium path results in alternative solution concepts to be used in implementation. The two incomplete information extensive form notions of implementation that have been analyzed are perfect Bayesian implementation (Brusco 1995, 1997, 1998) and implementation via sequential equilibrium (Baliga 1999 and Bergin and Sen 1998). The interesting new aspect that arises in these settings is that preference reversals can arise not only

[54] See Dutta and Sen (1991c) for more on the necessary and sufficient conditions for Bayesian implementation.

[55] See also Matsushima (1993).

[56] Other natural extensions of complete information implementation to incomplete information settings have been analyzed as well, such as implementation in undominated Bayesian equilibrium (Palfrey and Srivastava 1989), virtual Bayesian implementation (Duggan 1997a), and virtual implementation in iteratively undominated strategies with incomplete information (Abreu and Matsushima 1992).

because of differences in the primitive underlying preferences, but also from the way in which information is revealed in equilibrium through the extensive form. This is the focus of the work of Bergin and Sen (1998). While various conditions for implementation have been identified, the role of information revelation through an extensive form is not yet fully understood.[57]

7.2 Ex post individual rationality, renegotiation, and credibility

At several points I have mentioned that various forms of implementation rely on the belief that the outcomes of the mechanism will be enforced, even if they are "bad" from society's point of view ex-post. This can be problematic, to the extent that the positive results depend on such outcomes being used by the mechanism and such beliefs holding.[58] If, for example, a mechanism is constructed to assist bargainers in reaching an efficient agreement, then it is questionable to assume that highly inefficient outcomes will be allowed to stand off (or on) the equilibrium path.

There are several papers that consider implementation in the face of individual rationality and/or renegotiation on and off the equilibrium path. Ma et al. (1988) were the first to point out the importance of imposing an individual rationality constraint both in and out of equilibrium. They examined a principal-agent model where the usual individual rationality constraint (imposed only on the equilibrium path) was replaced by an "opt-out," where each player had the ability to decline the outcome of the mechanism and accept a status-quo outcome instead. Maskin and Moore (1999) examined a more general implementation problem, and changed the opting out to a possibility of renegotiation. This takes the point of view that instead of settling for some outside option or status-quo, the individuals involved are likely to renegotiate to some efficient outcome. Maskin and Moore considered implementation where any outcome of a mechanism that suggests a Pareto dominated allocation is replaced by a Pareto efficient allocation according to an exogenous renegotiation function.[59] Given any such renegotiation function, Maskin and Moore obtain characterizations of Nash and subgame perfect implementation that have intuitive relationships to the standard characterizations. Although these two papers provide insight into ex-post individual rationality and renegotiation, much depends on the exogenous specification of the outside options or renegotiation function.

One would like to model the process that occurs when players opt-out, rather than take as given a status quo or renegotiation function. Jackson and Palfrey (1998) push in this direction, in the context of a dynamic bargain-

[57] For instance, see Brusco (1998) for a puzzle on the necessity of multi-stage mechanisms.

[58] See Hurwicz (1994) for a discussion of some other, related issues related to enforceability in mechanism design.

[59] See Rubinstein and Wolinsky (1992) for a different approach. Their notion of "renegotiation-proof" implementation requires that the possibility for renegotiation never arise on or off the equilibrium path.

ing and matching model,[60] by having a player be rematched with a new bargaining partner whenever he or she opts-out of a prescribed alternative. They show that although such an endogenous individual rationality constraint is compatible with efficiency within individual matches, it can be incompatible with efficiency from society's point of view accounting for the overall evolution of the market which requires specific exercise of the rematching option. In other words, it can be impossible to implement the efficient rule in such a setting.

The above approaches are based on the idea that the individuals themselves are not bound to the mechanism, but have the ability to opt out of the prescribed outcome either to some status quo, renegotiated outcome, or replay of the mechanism. There might be other contexts where such an opt-out can be prevented and the outcome can be made legally binding. Nevertheless, one still has to worry about whether the planner (or society at large) will let inefficient outcomes stand ex-post. Studies which address such "credibility", or the ability of the mechanism designer or planner to commit to off-equilibrium-path outcomes that are known to be undesirable, include Chakravorti et al. (1992), Baliga et al. (1995) and Baliga and Sjöström (1995). These studies include the planner as a player in the mechanism,[61] in which case one can explicitly account for the planner's preferences and behavior with regards to enforcing an outcome.

In all of the above work there are two forces at work. On the one hand, allowing for movement away from ex-post undesirable outcomes can be improving just by itself since truly undesirable outcomes are eliminated automatically. On the other hand, this limits out of equilibrium threats that (as we saw in the last section) can play a strong role in selecting which outcomes are implemented. Thus, although allowing for players or the planner to alter outcomes builds in some minimal individual rationality or efficiency, it can come at the expense of selectivity of the mechanism. While the above cited work makes progress in understanding enforceability and credibility issues, a complete understanding of these issues requires full modeling (endogeneity) of the process that occurs when an outcome is opted away from by the players or the planner. Often this process may be situation specific, and so it should be worthwhile to analyze this issue in a variety of specific environments, which is something this branch of the literature is just beginning to do.

7.3 Mixed strategies

We do not yet have a good understanding of what is Nash or subgame perfect implementable when one accounts for mixed strategies, by mechanisms satisfying, for instance, the best response property.[62] I discussed one approach

[60] See Jackson and Palfrey (2001) for a unified approach to dealing with renegotiation, outside options, and replay of the mechanism, in more abstract settings, with some applications to exchange economies.

[61] Including the planner as a player has an interesting theoretical byproduct: the planner is part of the equilibrium and can thus rule out undesired equilibria unilaterally.

[62] As mentioned in the discussion following Example 5, there are mechanisms that have no mixed strategy equilibria; but those mechanisms involve integer games and fail to have well-defined best response correspondences.

to addressing this problem which is to analyze specific environments where the additional structure can lead to simple implementing mechanisms that avoid mixed strategy equilibria and other difficulties altogether. Such an approach has been very successful leading to some of the most natural mechanisms for implementation to date. So it seems fruitful to simply break the implementation analysis into many subproblems by analyzing specific settings and developing a catalog of what is implementable setting by setting. However, a case can still be made for attacking the problem head-on in the general setting. The reason for complementing the setting-specific analyses with a general analysis is that it would generate insight to the additional necessary (and sufficient) conditions that come with the consideration of mixed strategies. Given how tailored the analyses are in specific environments, it is not clear that such an understanding will be obtained by piecing together a catalog of results.

A number of possible head-on approaches suggest themselves, even simply with respect to Nash implementation. One might ask what can be (exactly) Nash implemented in finite environments by a finite mechanism that has no mixed strategies.[63] Alternatively, one could allow for mixed strategies and consider implementing social choice correspondences that map into lotteries and ask what is implementable.

7.4 Robustness of mechanisms

In order to push the theory towards applied use, one has to worry about how sensitive a mechanism is to such obvious things as misspecification of the environment or 'irrational' behavior by the players.

Some work has been done on implementation allowing for variations in the behavior of the individuals. Some of the first work in this direction was on the implementation of the Walrasian correspondence by mechanisms with a continuous outcome function (e.g., see Postlewaite and Wettstein 1989). Requiring continuity of a mechanism results in an outcome close to the desired one if some players deviate slightly from equilibrium strategies. More recent attention to robustness with respect to behavior appears in studies of double implementation (e.g., see Yamato 1993, 1994a,b). Double implementation requires implementation in more than one solution concept by the same mechanism. When allowing for combinations such as Nash and strong equilibrium, potentially unforeseen coalitional deviations are accounted for in implementation. One can also ask for robustness with regards to behavior that is completely unpredictable by some of the players. A new paper by Eliaz (1999) examines this sort of problem asking that a mechanism rely only on a majority of the individuals (not knowing ex-ante which majority it will be), and using a solution concept that requires behavior of the players to be immune to arbitrary behavior by some subsets of other players.

[63] Again, Abreu and Matsushima (1992a) provide a good deal of insight into this question for virtual implementation. The reason for considering the exact setting would be to loosen assumptions on preferences and sensitivity to small probability events under the lotteries.

While the work described above has examined robustness with regards to certain aspects of the behavior of players, there is almost no study of mechanisms that are immune to misspecifications of the environment. The only study in this direction that I am aware off is by Duggan and Roberts (1998), who study how continuity of a mechanism can result in robustness with respect to small errors in specification, due to corresponding upper-hemi continuity of equilibrium correspondences.[64] This suggests that an intuitive notion of requiring some continuity of a mechanism in order to guarantee some minimal robustness is well founded. Short of this, we are lacking even a basic understanding of the exact limitations that are imposed by requiring implementation by mechanisms that are robust to misspecifications of the domain of preferences, beliefs, allocations, or even the number or role of various individuals in a society.

7.5 Comparisons across solution concepts

As we have developed a deeper understanding of what is implementable in a variety of contexts and solution concepts, it has become clear that the choice of solution concept used to model behavior has an important bearing on what is implementable. The class of social choice correspondences that are implementable in undominated Nash or subgame perfect equilibrium is significantly larger than the class of Nash implementable social choice correspondences. What is it about solution concepts (or assumptions on individual behavior) that accounts for this difference? Jackson and Srivastava (1996) examine this issue in the context of a voting setting with a finite number of strictly ranked alternatives. They point to differences in consideration of off-equilibrium path behavior as the critical factor the affects how much is implementable via different solution concepts.[65] While this works out cleanly in a voting context, it is still unknown if the same is true more generally or even in other specific contexts like exchange economies.

Tracing the characterizations of implementability back to assumptions concerning individual behavior should be quite important. As modelers, we often consider a solution concept as an object of choice; but ultimately it boils down to understanding how people behave in a given context. If we understand how variations in behavior tie back to implementability, then we will eventually be well-equipped to make prescriptions in various contexts. We are still short of such an understanding.

7.6 Repeated and dynamic implementation

Perhaps one of the most interesting and relevant, and still least studied, questions in implementation theory is that of implementation in dynamic and re-

[64] This holds provided one worries about mixed strategies, which Duggan and Roberts point to as another reason for accounting for mixed strategies in implementation.

[65] New work by Kaya and Koray (2000) provides interesting characterizations based on identifying solutions that only implement monotonic social choice functions on a given class of mechanisms.

peated contexts. We know from results on folk theorems, reputation, and learning (as well as various experimental and applied studies), that behavior in repeated games can differ in important ways from behavior in one-shot games. As most applications of implementation are repeated ones, ranging from the design of voting procedure to markets, it is very important that we understand how mechanisms perform when used repeatedly by the same society. Work by Kalai and Ledyard (1998) is the only that I am aware of that takes this repeated perspective. They draw on intuition from the reputation and learning literatures to note that if a mechanism designer is more patient than the underlying population, then the society's characteristics can be teased out and learned over time. Thus in the long run, a planner can avoid even incentive compatibility requirements in a Bayesian setting.

There are also interesting contexts where a mechanism is not binding and so may be used an arbitrary number of times until an agreement is reached. This is the case in many negotiation and market settings, where a mechanism outlines potential transactions but a transaction is concluded only when all parties sign a binding agreement. This leads to a notion of implementation that is dynamic in nature, but quite different from repeated implementation as an outcome is only reached at one date which is endogenously determined. Such implementation is studied in Jackson and Palfrey (2001), where a notion of stationary implementation is examined that is roughly implementation in Markov perfect equilibria when agents may ask for a replay of the mechanism if they do not like the outcome it suggests. This notion of implementation has a relatively simple characterization, and allows for the implementation of the constrained Walrasian correspondence if agents discount the future.

While these papers take first steps towards modeling dynamic versions of implementation, the question of what is implementable in dynamic contexts is still largely open. Given that most design questions deal with institutions, such as markets or voting procedures, that will be used more than once, this is one of the most important open areas in the literature.

7.7 Endogenous choice of mechanisms

The theory of implementation is agnostic on how a social choice correspondence is selected for implementation or how a society might choose a mechanism. These are critical questions in understanding the eventual outcome of the process. These are also difficult questions to formulate, since at some point one has to make an ad hoc assumption about the process of choice. How do people choose how to choose? Nonetheless one can ask for some sort of consistency properties in decision making. Koray (2000), in a voting context, asks for consistency in that a social choice function chosen by a society should be consistent with the manner in which it is selected. That is, the selection of a social choice function can be formulated as a social choice problem, and the social choice function selected should be one that would be selected by itself. A clever formulation (see Koray 2000) allows social choice functions to be both the objects of choice and the methods of choice at the same time. Although Koray's result is negative (he finds that only a dictatorial social choice function

is self-consistent), one can think of addressing such issues on restricted domains and allowing for correspondences rather than functions.[66]

The question of endogenous determination of mechanisms is one of particular importance if we wish to push implementation theory to the full extent of positive implications and not simply normative ones.[67] This is yet another area with largely unexplored questions and themes.

7.8 Mechanisms in a larger natural context

Having a useful theory of implementation requires not only that we develop an understanding of implementation via mechanisms with a variety of nice properties, but also that we keep track of the fact that the mechanism will generally sit in a larger "natural" context that the designer may have no control over. This issue is raised by Hurwicz (1994) where he points out that individuals may have strategies that are beyond the control of the designer.[68] Here, almost by definition, the issues will be context specific and so we should not expect to develop general abstract models that encompass large sets of applications. Instead the variety of issues that arise will be specific to the problem at hand.

One issue that has been looked at is endogenizing the set of alternatives in economic settings. This was explored by Postlewaite (1979) (see also Thomson 1977) who examined the possibility of individuals withholding (or destroying) part of their endowments, which can lead to improvements for individuals. Systematic analyses of endogenous endowments or production possibilities in implementation in economic settings have been provided by Hurwicz et al. (1995) and Hong (1996, 1998).

Settings other than production and exchange economies also have alternatives that are to some extent endogenous beyond the control of the mechanism. This endogeneity can be critical to evaluating differences in mechanisms as different mechanisms may provide very different incentives with regards to how individuals behave in the larger context, and thus to which outcomes are eventually realized. Other work in this spirit includes studies where the set of candidates in a voting context are endogenous as explored in Dutta et al. (1998, 2001), Eraslan and McLennan (2000), and Rodriguez-Alvarez (2000), where mechanisms may be selected as in Lagunoff (1992), and situations where

[66] New work by Barbera and Jackson (2000) avoids this impossibility result by focussing on settings where voting is over two alternatives – a status quo and an alternative. Voting rules are simply characterized by a quota of how many supporters are needed to have the status-quo defeated. Under simple assumptions on uncertainty, voters individuals have nicely behaved preference over voting procedures. Barbera and Jackson examine the existence of self-stable voting procedures: a given voting procedure is self-stable if it is not defeated by some other voting procedure when the given voting procedure is used to make the choice.

[67] A related branch of the mechanism design literature relates to the choice of mechanisms by competing sellers, as first studied by McAfee (1993).

[68] This can also be related back to the issue of credibility, renegotiation, and ex post individual rationality as discussed above, and so that discussion could very well have been framed in this section.

there are endogenous transfer payments from one player to another that may affect the equilibrium structure of a mechanism, as in Jackson and Wilkie (2000). The reason that this is an important issue is that the predictions of what outcomes we might expect can be quite different if the individuals have strategic choices (like deciding whether to be a candidate, or bribing other individuals to take certain actions) that are outside of the standard implementation analysis. If such actions are available and unaccounted for then the standard analysis that ignores such outside actions could be quite misleading.

A related matter is that an individual's preferences and behavior may depend on the mechanism itself. Glazer and Rubinstein (1998) present a context where individuals providing messages ("opinions") care explicitly about how their message ends up comparing to the ultimate social decision, as they would like to have their recommendation match the social choice. This introduces an added layer of feedback between the mechanism and equilibrium behavior, that presents interesting and challenging issues for implementation theory. Such issues are generally unexplored in the theory, and one can think of any number of examples beyond the Glazer and Rubinstein context (for instance in principal-agent problems) where they are important.

7.9 Testing implementing mechanisms

As the theory continues to generate mechanisms for implementation, we can begin to evaluate and compare them through testing, simulation, experiments, and ultimately empirical work. For instance Chen (1997) compares behavior observed in experiments involving a series of public goods mechanisms. She attributes differences in performance to supermodularity properties of the mechanisms, which then helps us appreciate another feature that we might consider in designing mechanisms. One can also test some of the canonical mechanisms directly as done by Sefton and Yavas (1996). They ran experiments on the Abreu and Matsushima (1992a) mechanism, and present some formidable hurdles for the theory.[69]

Simulations can be also be used to study the same issues, as in Cabrales (1999). He studies adaptive learning by players in some of the canonical mechanisms for implementation. The adaptive learning algorithm Cabrales analyzes has features that allow for complete randomization over best responses which can kick the process out of an integer game, and thus produce convergence to equilibrium in Nash mechanisms. Ultimately, one would like a mechanism to perform well under a variety of processes, and to offer fast convergence to equilibrium as players learn or evolve.

[69] These just touch the surface of what can be explored, and one can check whether various criticisms of unnatural features of mechanisms are supported by experimental evidence. For instance, how a mechanism with an integer game would be played? Would players play equilibrium strategies? Would they end up trying to announce high integers? (I am sure that Caltech undergraduates could be very creative in announcing high integers.) How would behavior depend on the social choice correspondence being implemented? Are they more likely to coordinate on the equilibrium play when there is alignment of preferences?

There is ample room for tests, simulations, experiments, and empirical research on mechanisms to add valuable guidance and insight to the theory of implementation. Issues such as how well various mechanisms perform when players are not at equilibrium but learning or adjusting according to some dynamic are quite important and some of the first issues that come to mind when one takes an experimental point of view; and yet have not even been touched by implementation theory.[70]

References

Abreu D, Matsushima H (1992a) Virtual implementation in iteratively undominated strategies I: Complete information. Econometrica 60: 993–1008

Abreu D, Matsushima H (1992b) Virtual implementation in iteratively undominated strategies II: Incomplete information. mimeo

Abreu D, Matsushima H (1992c) A response to Glazer and Rosenthal. Econometrica 60: 1439–1442

Abreu D, Matsushima H (1994) Exact implementation. J Econ Theory 64: 1–19

Abreu D, Sen A (1991) Virtual implementation in Nash equilibria. Econometrica 59: 997–1022

Abreu D, Sen A (1990) Subgame perfect implementation: A necessary and almost sufficient condition. J Econ Theory 50: 285–299

Allen B (1997) Implementation theory with incomplete information. In: Hart S, Mas-Colell A, Cooperation: Game theoretic approaches. Springer, Berlin Heidelberg New York

Amarós P, Moreno B (2001) Implementation of optimal contracts under adverse selection. Rev Econ Design 6: 41–62

Arya A, Glover J (1995) A simple forecasting mechanism for moral hazard settings. J Econ Theory 66: 507–521

Arya A, Glover J, Hughes J (1997) Implementing coordinated team play. J Econ Theory 74: 218–232

Arya A, Glover J, Rajan U (2000) Implementation in principal-agent models of adverse selection. J Econ Theory 93: 87–109

Arya A, Glover J, Young R (1995) Virtual implementation separable Bayesian environments using simple mechanisms. Games Econ Behav 9: 127–138

Bag PK (1996) Efficient allocation of a ''pie'': King Solomon's dilemma. Games Econ Behav 12: 21–41

Bag PK (1997) Public goods provision: Applying Jackson-Moulin mechanisms for restricted agent characteristics. J Econ Theory 73: 460–472

Bagnoli M, Lipman B (1989) Provision of public goods: Fully implementing the core through private contributions. Rev Econ Stud 56: 583–601

Baliga S (1999) Implementation in economic environments with incomplete information: The use of multi-stage games. Games Econ Behav 27: 173–183

Baliga S, Brusco S (2000) Collusion, renegotiation, and implementation. Soc Choice Welfare 17: 69–84

Baliga S, Corchon L, Sjöström T (1997) The theory of implementation when the planner is a player. J Econ Theory 77: 15–33

[70] There is a rich literature in game theory on this topic, but it has not been looked at from the perspective of designing mechanism to have nice learning or dynamic properties.

Baliga S, Sjöström T (1998) Decentralization and collusion. J Econ Theory 83: 196–232
Baliga S, Sjöström T (1999) Interactive implementation. Games Econ Behav 27: 38–63
Bandyopadhyay T, Samuelson L (1992) Weakly implementable social choice rules. Theory Decision 33: 135–151
Barberà S (2001) An introduction to strategy-proof social choice functions. Soc Choice Welfare 18: 619–653
Barberà S, Dutta B (1982) Implementation via protective equilibria. J Math Econ 4: 49–65
Barberà S, Jackson MO (2000) Choosing how to choose: Self-stable majority rules. mimeo, Caltech
Bergin J (1993) On some recent results in incomplete information implementation. mimeo, Queen's University
Bergin J, Sen A (1996) Implementation in generic environments. Soc Choice Welfare 13: 467–478
Bergin J, Sen A (1998) Extensive form implementation in incomplete information environments. J Econ Theory 80: 222–256
Bergin J, Duggan J (1999) An implementation theoretic approach to non-cooperative foundations. J Econ Theory 86: 50–76
Bernheim D, Whinston M (1987) Coalition-proof Nash equilibrium II: Applications. J Econ Theory 42: 13–29
Blume L, Easley D (1990) Implementation of Walrasian expectations equilibria. J Econ Theory 51: 207–227
Boylan R (1998) Coalition-proof implementation. J Econ Theory 82: 132–143
Brusco S (1995) Perfect Bayesian implementation. Econ Theory 5: 419–444
Brusco S (1997) Implementing action profiles when agents collude. J Econ Theory 73: 395–424
Brusco S (1999) Implementation with extensive form games: One round of signaling is not enough. J Econ Theory 87: 356–378
Brusco S (1998a) Implementing action profiles with sequential mechanisms. Rev Econ Design 3: 271–300
Brusco S (1998b) Unique implementation of the full surplus extraction outcome in auctions with correlated types. J Econ Theory 80: 185–200
Brusco S (1998c) Perfect Bayesian implementation in economic environments. mimeo, Universidad Carlos III de Madrid
Brusco S, Jackson M (1999) The optimal design of a market. J Econ Theory 88: 1–39
Cabrales A (1999) Adaptive dynamics and the implementation problem with complete information. J Econ Theory 86: 159–184
Chakravorti B (1991) Strategy space reductions for feasible implementation of Walrasian performance. Soc Choice Welfare 8: 235–246
Chakravorti B, Corchon L, Wilkie S (1992) Credible implementation. Games Econ Behav (forthcoming)
Chattopadhyay S, Corchon L, Naeve J (2000) Contingent commodities and implementation. Econ Letters 68: 293–298
Chen Y (1997) Dynamic stability of Nash-efficient public goods mechanisms: reconciling theory and experiments. mimeo, University of Michigan
Corchon L (1996) The theory of implementation of socially optimal decisions in economics. McMillan, London
Corchon L, Herrero C (1995) A decent proposal. mimeo, Universidad de Alicante
Corchon L, Ortuno-Ortin I (1995) Robust implementation under alternative information structures. Econ Design 1: 159–172
Corchon L, Wilkie S (1996) Double implementation of the ratio correspondence by a market mechanism. Econ Design 2: 325–337
Dagan N, Serrano R, Volij O (1999) Feasible implementation of taxation methods. Rev Econ Design 4: 57–72

Danilov V (1992) Implementation via Nash equilibrium. Econometrica 60: 43–56
Dasgupta P, Hammond P, Maskin E (1979) The implementation of social choice rules: Some general results on incentive compatibility. Rev Econ Stud 46: 185–216
Dasgupta P, Maskin E (1997) Notes on efficient auctions. mimeo, Harvard University
Demange G (1984) Implementing efficient egalitarian equivalent allocations. Econometrica 52: 1167–1178
Demski J, Sappington D (1984) Optimal incentive contracts with multiple agents. J Econ Theory 33: 152–171
Duggan J (1995) Sequentially rational implementation with incomplete information. mimeo, University of Rochester
Duggan J (1997a) Virtual Bayesian implementation. Econometrica 65: 1175–1199
Duggan J (1997b) Implementing social welfare optima when there is a private good. mimeo, University of Rochester
Duggan J (1998) An extensive form solution to the adverse selection problem in principal/multi-agent environments. Rev Econ Design 3: 167–191
Duggan J, Roberts J (1997) Robust implementation. mimeo, University of Rochester
Dutta B, Jackson MO, Le Breton M (2001) Strategic candidacy and voting procedures. Econometrica 69: 1013–1038
Dutta B, Jackson MO, Le Breton M (1998) Voting by successive elimination and strategic candidacy. J Econ Theory (forthcoming)
Dutta B, Sen A (1991a) Implementation under strong equilibria: A complete characterization. J Math Econ 20: 49–68
Dutta B, Sen A (1991b) Necessary and sufficient conditions for 2-person Nash implementation. Rev Econ Stud 58: 121–129
Dutta B, Sen A (1991c) Implementation in Bayesian equilibrium. mimeo, Indian Statistical Institute
Dutta B, Sen A (1993) Implementing generalized condorcet social choice functions via backward induction. Soc Choice Welfare 10: 149–160
Dutta B, Sen A (1994) The necessity of infinite mechanisms. J Econ Theory 64: 130–141
Dutta B, Sen A, Vohra R (1995) Nash implementation through elementary mechanisms in economic environments. Econ Design 1: 173–204
Ehlers L (2000) Monotonic and implementable solutions in generalized matching problems. mimeo, Maastricht University
Eliaz K (1999) Fault tolerant implementation. mimeo, Tel Aviv University
Eraslan H, McLennan A (2000) Strategic candidacy for multivalued voting procedures. mimeo, University of Minnesota
Gevers L (1986) Walrasian social choice: Some simple axiomatic approaches. In: Heller W et al., Social choice and public decision making: Essays in honor of Kenneth J. Arrow. Cambridge University Press, Cambridge
Gibbard A (1973) Manipulation of voting schemes: A general result. Econometrica 41: 587–601
Giraud G, Rochon C (2001) Consistent collusion-proofness and correlation in exchange economies. mimeo, CORE
Glazer J, Ma A (1989) Efficient allocation of a "Prize": King Solomon's dilemma. Games Econ Behav 1: 222–233
Glazer J, Perry M (1996) Virtual implementation in backwards induction. Games Econ Behav 15: 27–32
Glazer J, Rosenthal R (1992) A note on Abreu–Matsushima mechanisms. Econometrica 60: 1435–1438
Glazer J, Rubinstein A (1998) Motives and implementation: On the design of mechanisms to elicit opinions. J Econ Theory 79: 157–173
Glover J (1994) A simpler mechanism that stops agents from cheating. J Econ Theory 62: 221–229

Groves T (1977) Efficient collective choice with compensation. In: Laffont J-J, Aggregation and revelation of preferences. North-Holland, Amsterdam
Hahn G, Yannelis N (1995) Coalitional Bayesian Nash implementation in differential information economies. mimeo, University of Illinois
Harsanyi JC (1967–68) Games with incomplete information played by 'Bayesian' Players. Management Science 14: 159–189, 320–334, 486–502
Herrero M, Srivastava S (1992) Implementation via backward induction. J Econ Theory 56: 70–88
Hong L (1995) Nash implementation in production economies. J Econ Theory 5: 401–417
Hong L (1996) Bayesian implementation in exchange economies with state dependent feasible sets and private information. Soc Choice Welfare 13: 433–444
Hong L (1998) Feasible Bayesian implementation with state dependent feasible sets. J Econ Theory 80: 201–221
Hong L, Page S (1994) Reducing informational costs in endowment mechanisms. Econ Design 1: 103–117
Hurwicz L (1972) On informationally decentralized systems. In: McGuire CB, Radner R, Decision and organization. North Holland, Amsterdam
Hurwicz L (1973) The design of mechanisms for resource allocation. Am Econ Rev 61: 1–30
Hurwicz L (1977) On allocations attainable through Nash equilibria. In: Laffont J-J, Aggregation and revelation of preferences. North-Holland, Amsterdam
Hurwicz L (1979) Outcome functions yielding Walrasian and Lindahl allocations at Nash equilibrium points. Rev Econ Stud 46: 217–225
Hurwicz L (1994) Economic design, adjustment processes, mechanisms and institutions. Econ Design 1: 1–14
Hurwicz L (1996) Feasible balanced outcome functions yielding constrained Walrasian and Lindahl allocations at Nash equilibrium points in economies with two agents when the designer knows the feasible set. mimeo, University of Minnesota
Hurwicz L, Schmeidler D (1978) Construction of outcome functions guaranteeing existence and pareto optimality of Nash equilibria. Econometrica 46: 1447–1474
Hurwicz L, Maskin E, Postlewaite A (1995) Feasible Nash implementation of social choice correspondences when the designer does not know endowments or production sets. In: Ledyard J, The economics of information decentralization: Complexity, efficiency, and stability. Kluwer, Amsterdam
Jackson MO (1991) Bayesian implementation. Econometrica 59: 461–478
Jackson MO (1992) Implementation in undominated strategies: A look at bounded mechanisms. Rev Econ Stud 59: 757–775
Jackson MO, Moulin H (1992) Implementing a public project and distributing its cost. J Econ Theory 57: 125–140
Jackson MO, Palfrey T (1998) Efficiency and voluntary implementation in markets with repeated pairwise bargaining. Econometrica 66: 1353–1388
Jackson MO, Palfrey T (2001) Voluntary implementation. J Econ Theory 98: 1–25
Jackson MO, Palfrey T, Srivastava S (1994) Undominated Nash implementation in bounded mechanisms. Games Econ Behav 6: 474–501
Jackson MO, Srivastava S (1992) On two person Nash implementable social choice functions. Soc Choice Welfare 9: 263–264
Jackson MO, Srivastava S (1996) Characterizations of game theoretic solution concepts which lead to impossibility theorems. Rev Econ Stud 63: 23–38
Jackson MO, Wilkie S (2000) Endogenous games and mechanisms: Side payments among players. mimeo, Caltech
Kalai E, Ledyard J (1998) Repeated implementation. J Econ Theory 83: 308–317
Kara T, Sönmez T (1996) Nash implementation of matching rules. J Econ Theory 68: 425–439

Kara T, Sönmez T (1997) Implementation of college admissions rules. Econ Theory 9: 197–218
Kaya A, Koray S (2000) Two essays in social choice theory. mimeo, Bilkent University
Koray S (2000) Self-selective social choice functions verify arrow and Gibbard-Satterthwaite theorems. Econometrica 68: 981–996
Kwasnica AM (1998) Bayesian implementable efficient and core allocations. mimeo, Caltech
Lagunoff R (1992) Fully endogenous mechanism selection on finite outcome sets. Econ Theory 2: 462–480
Li Q, Nakamura S, Tian G (1995) Nash implementation of the Lindahl correspondence with decreasing returns to scale technologies. Int Econ Rev 36: 37–52
Ma C (1988) Unique implementation of incentive contracts with many agents. Rev Econ Stud 55: 555–572
Ma C, Moore J, Turnbull S (1988) Stopping agents from 'cheating'. J Econ Theory 46: 355–372
Maniquet F (1994) On equity and implementation in economic environments. Ph.D. dissertation, University of Namur
Maniquet F (1999) Implementation under perfect information in economic environments. mimeo, University of Namur
Mas–Colell A, Vives X (1993) Implementation in economies with a continuum of agents. Rev Econ Stud 60: 613–629
Maskin E (1985) The theory of implementation in Nash equilibrium. In: Hurwicz L, Schmeidler D, Sonnenschein H (eds) Social goals and social organization: Essays in honor of Elisha A. Pazner. Cambridge University Press, Cambridge
Maskin E (1999) Nash equilibrium and welfare optimality. Rev Econ Stud 66: 23–38
Maskin E, Moore J (1999) Implementation with renegotiation. Rev Econ Stud 66: 39–56
Maskin E, Sjöström T (2001) Implementation theory. mimeo, Harvard University and Penn State
Matsushima H (1988) A new approach to the implementation problem. J Econ Theory 45: 128–144
Matsushima H (1993) Bayesian monotonicity with side payments. J Econ Theory 59: 107–121
McAfee RP (1993) Mechanism design by competing sellers. Econometrica 61: 1281–1312
McKelvey R (1989) Game forms for Nash implementation of general social choice correspondences. Soc Choice Welfare 6: 139–156
Miyagawa E (1997) Implementation of the normalized utilitarian bargaining solution. mimeo, University of Rochester
Miyamoto H, Watanabe T, Mizuno S (1990) Implementation in admissible strategies. Technical Report No. 28, Department of Management Science and Engineering, Tokyo Institute of Technology
Mookherjee D, Reichelstein S (1990) Implementation via augmented revelation mechanisms. Rev Econ Stud 57: 453–476
Mookherjee D, Reichelstein S (1992) Dominant strategy implementation of Bayesian incentive compatible allocation rules. J Econ Theory 56: 378–399
Moore J (1992) Implementation, contracts, and renegotiation environments with complete information. In: Laffont J-J, Advances in economic theory. Cambridge University Press, Cambridge
Moore J, Repullo R (1988) Subgame perfect implementation. Econometrica 56: 1191–1220
Moore J, Repullo R (1990) Nash implementation: A full characterization. Econometrica 58: 1083–1100
Moulin H (1980b) Implementing efficient, anonymous and neutral social choice functions. J Math Econ 72: 249–269

Moulin H (1981) Implementing just and efficient decision making. J Publ Econ 16: 193–213
Moulin H (1982) Non-cooperative implementation: A survey of recent results. Math Soc Sci 3: 243–257
Moulin H (1988) Axioms of cooperative decision making. Cambridge University Press, Cambridge
Mount K, Reiter S (1974) The informational size of message spaces. J Econ Theory 8: 161–192
Muller E, Satterthwaite M (1977) On the equivalence of strong positive association and strategy–proofness. J Econ Theory 14: 412–418
Myerson R, Satterthwaite M (1983) Efficient mechanisms for bilateral trading. J Econ Theory 29: 265–281
Nagahisa RI (1994) A necessary and sufficient condition for Walrasian social choice. J Econ Theory 62: 186–208
Nakamura S (1988) Feasible Nash implementation of competitive equilibria in an economy with externalities. mimeo, University of Minnesota
Nakamura S (1990) A feasible Nash implementation of Walrasian equilibria in the two agent economy. Econ Lett 34: 5–9
Nakamura S (1998) Impossibility of Nash implementation in two person economies. Econ Design 3: 159–166
Osana H (1997) Nash-implementation of the weak pareto choice rule for indecomposable environments. Rev Econ Design 3: 57–74
Osana H (1990) Implementation of multi-agent incentive contracts with the principal's renegotiation offer. Rev Econ Design 4: 161–178
Palfrey T (1990) Implementation in Bayesian equilibrium: The multiple equilibrium problem in mechanism design. In: Laffont J-J, Advances in economic theory. Cambridge University Press, Cambridge
Palfrey T (1995) Implementation theory. In: Aumann R, Hart S, Handbook of game theory (forthcoming)
Palfrey T, Srivastava S (1986) Private information in large economies. J Econ Theory 39: 34–58
Palfrey T, Srivastava S (1987) On Bayesian implementable allocations. Rev Econ Stud 54: 193–208
Palfrey T, Srivastava S (1989a) Implementation with incomplete information in exchange economies. Econometrica 57: 115–134
Palfrey T, Srivastava S (1989b) Mechanism design with incomplete information: A solution to the implementation problem. J Polit Econ 97: 668–691
Palfrey T, Srivastava S (1991) Nash implementation using undominated strategies. Econometrica 59: 479–502
Palfrey T, Srivastava S (1993) Bayesian implementation. Harwood Academic Publishers, Switzerland
Peleg B (1996a) A continuous double implementation of the constrained Walrasian equilibrium. Econ Design 2: 89–97
Perry M, Reny P (1999) A general solution to King Solomon's dilemma. Games Econ Behav 26: 279–285
Peleg B (1996b) Double implementation of the Lindahl equilibrium by a continuous mechanism. Econ Design 2: 311–324
Piketty T (1993) Implementation of first-best allocations via generalized tax schedules. J Econ Theory 61: 23–41
Postlewaite A (1985) Implementation via Nash equilibrium in economic environments. In: Hurwicz L, Schmeidler D, Sonnenschein H (eds) Social goals and social organization: Essays in honor of Elisha A. Pazner. Cambridge University Press, Cambridge
Postlewaite A (1979) Manipulation via endowments. Rev Econ Stud 46: 255–262
Postlewaite A, Schmeidler D (1986) Implementation in differential information economies. J Econ Theory 39: 14–33

Postlewaite A, Wettstein D (1989) Continuous and feasible implementation. Rev Econ Stud 56: 14–33
Rai A (1997) Efficient audit mechanisms to target the poor. Dissertation, University of Chicago
Reichelstein S, Reiter S (1988) Game forms with minimal strategy spaces. Econometrica 56: 661–692
Repullo R (1987) A simple proof of Maskin's theorem on Nash implementation. Soc Choice Welfare 4: 39–42
Rochon C (2000) Stability and efficiency: Monetary policy and implementation theory. Ph.D. Dissertation, Université Catholique de Louvain
Rodriguez-Alvarez C (2000) Candidate stability and multi-valued voting procedures. mimeo, Universitat Autonoma de Barcelona
Rubinstein A, Wolinsky A (1992) Renegotiation-proof implementation and time preferences. Am Econ Rev 82: 600–614
Saglam I (1997) A note on Jackson's theorems in Bayesian implementation. mimeo, Bilkent University
Saijo T (1987) On constant Maskin-monotonic social choice functions. J Econ Theory 42: 382–386
Saijo T (1988) Strategy space reduction in Maskin's theorem: Sufficient conditions for Nash implementation. Econometrica 56: 693–700
Saijo T, Tatimatani Y, Yamato T (1996) Toward natural implementation. Int Econ Rev 37: 949–980
Schmeidler D (1980) Walrasian analysis via strategic outcome functions. Econometrica 48: 1585–1593
Sefton M, Yavas A (1996) Abreu-Matsushima mechanisms: Experimental evidence. Games Econ Behav 16: 280–302
Sen A (1995) The implementation of social choice functions via social choice correspondences: A general formulation and a limit result. Soc Choice Welfare 12: 277–292
Serrano R (1993) Non-cooperative implementation of the nucleolus- the 3-player case. Int J Game Theory 22: 345–357
Serrano R (1995) A market to implement the core. J Econ Theory 67: 285–294
Serrano R (1997) A comment on the Nash program and the theory of implementation. Econ Lett 55: 203–208
Serrano R, Vohra R (1997) Non-cooperative implementation of the core. Soc Choice Welfare 14: 513–525
Serrano R, Vohra R (2001) Some limitations of virtual Bayesian implementation. Econometrica 69: 785–792
Serrano R, Vohra R (2000) Type diversity and virtual Bayesian implementation. mimeo, Brown University
Sertel M (1994) Manipulating Lindahl equilibrium via endowments. Econ Lett 46: 167–171
Shin S, Suh S (1996) A mechanism implementing the stable rule in marriage problems. Econ Lett 51: 169–176
Shin S, Suh S (1997) Double implementation by a simple game form in the commons problem. J Econ Theory 77: 205–213
Sjöström T (1991) On the necessary and sufficient conditions for Nash implementation. Soc Choice Welfare 8: 333–340
Sjöström T (1993) Implementation in perfect equilibrium. Soc Choice Welfare 10: 97–106
Sjöström T (1994) Implementation in undominated Nash equilibria without integer games. Games Econ Behav 6: 502–511
Sjöström T (1995a) Implementation in teams. Econ Design 1: 327–342
Sjöström T (1995b) Implementation by demand mechanisms. Econ Design 1: 343–354

Sjöström T (1996) Credibility and renegotiation of outcome functions in implementation. Japanese Econ Rev 47: 157–169
Sjöström T (1999) Undominated Nash implementation with collusion and renegotiation. Games Econ Behav 26: 337–352
Sönmez T (1996) Implementation in generalized matching problems. J Math Econ 26: 169–176
Suh S-C (1995) A mechanism implementing the proportional solution. Econ Design 1: 301–317
Suh S-C (1996) Implementation with coalition formation: A complete characterization. J Math Econ 25: 109–122
Suh S-C (1997a) An algorithm checking strong Nash implementability. J Math Econ 25: 109–122
Suh S-C (1997b) Games implementing the stable rule of marriage problems in strong Nash equilibria. mimeo, University of Windsor
Suh S-C (2001) An algorithm for verifying the double implementability in Nash and strong Nash equilibria. Math Soc Sci 41: 103–110
Tadenuma K, Toda M (1998) Implementable stable solutions to pure matching problems. Math Soc Sci 35: 121–132
Tatamitani Y (1991) Double implementation in Nash and undominated Nash equilibria in social choice environments. mimeo, University of Tsukuba
Thomson W (1977) Comment on: L. Hurwicz – On allocations attainable through Nash equilibria. In: Laffont J-J, Aggregation and revelation of preferences. North-Holland, Amsterdam
Thomson W (1979) Maximin strategies and elicitation of preferences. In: Laffont J-J, Aggregation and revelation of preferences. North Holland, Amsterdam
Thomson W (1996) Concepts of implementation. Japanese Econ Rev 47: 133–143
Thomson W (1994) Divide and permute and the implementation of solutions to the problem of fair division. mimeo, University of Rochester
Thomson W (1999) Monotonic extensions on economic domains. Rev Econ Design 4: 13–34
Tian G (1989) Implementation of the Lindahl correspondence by a single-valued, feasible, and continuous mechanism. Rev Econ Stud 56: 613–621
Tian G (1990) Completely feasible continuous implementation of the Lindahl correspondence with a message space of minimal dimensions. J Econ Theory 51: 443–452
Tian G (1991) Implementation of the Lindahl allocations with nontotal-nontransitive preferences. J Publ Econ 46: 247–259
Tian G (1993) Implementing Lindahl allocations by a withholding mechanism. J Math Econ 22: 169–179
Tian G (1994) Implementation of linear cost share equilibrium allocations. J Econ Theory 64: 568–584
Tian G (1996) Continuous and feasible implementation of rational expectations Lindahl allocations. Games Econ Behav 16: 135–151
Tian G (1999) Bayesian implementation in exchange economies with state dependent preferences and feasible sets. Soc Choice Welfare 16: 99–120
Tian G, Li Q (1991) Completely feasible and continuous implementation of the Lindahl correspondence with any number of goods. Math Soc Sci 21: 67–79
Tian G, Li Q (1994) An implementable and informational efficient state-ownership system with variable returns. J Econ Theory 64: 286–297
Tian G, Li Q (1995) On Nash-implementation in the presence of withholding. Games Econ Behav 9: 222–233
Trockel W (1998) An exact implementation of the Nash bargaining solution in dominant strategies. In: Abramovich Y, Avgerinos E, Yanellis N (eds) Functional analysis and economic theory. Springer, Berlin Heidelberg New York
van Damme E (1987) Stability and perfection of Nash equilibria. Springer, Berlin Heidelberg New York

Vartiainen H (2000) Subgame perfect implementation: A full characterization for the many-player case. mimeo, University of Helsinki
Vartiainen H (2000b) Subgame perfect implementation: A full characterization for the two-player case. mimeo, University of Helsinki
Wettstein D (1990) Continuous implementation of constrained rational expectations equilibria. J Econ Theory 52: 208–222
Wettstein D (1992) Continuous implementation in economies with incomplete information. Games Econ Behav 4: 463–483
Williams S (1984) Sufficient conditions for Nash implementation. mimeo, Northwestern University
Williams S (1986) Realization and Nash implementation: Two aspects of mechanism design. Econometrica 54: 139–151
Yamato T (1992) On Nash implementation of social choice correspondences. Games Econ Behav 4: 484–492
Yamato T (1993) Double implementation in Nash and undominated Nash equilibria. J Econ Theory 59: 311–323
Yamato T (1994) Equivalence of Nash implementability and robust implementability with incomplete information. Soc Choice Welfare 11: 289–303
Yamato T (1999) Nash implementability and double implementation: Equivalence theorems. J Math Econ 31: 215–238
Yoshihara N (1999) Natural and double implementation of public ownership solutions in differentiable production economies. Rev Econ Design 4: 127–152
Yoshihara N (2000) A characterization of natural and double implementation in production economies. Soc Choice Welfare 17: 571–600