Chapter 15

NASH IMPLEMENTATION

1. Introduction

In this chapter we continue to explore issues of manipulation of SCF's. In the last one we stumbled with an important impossibility result, the Gibbard-Satterthwaite theorem. Recall that this theorem establishes that, when there are at least three alternatives, there is no SCF that is nontrivial, universal, nonmanipulable and nondictatorial.

Given this negative result, one is forced to relax some of its assumptions in the hope of finding more encouraging news. For example, if society is choosing between only two alternatives, one can easily find SCF's that satisfy all these properties (such as majority voting). A second way out is based on the relaxation of universality: if the Central Authority has determined that certain kinds of preferences can be ruled out, possibility results arise. For example, if we assume that preferences are additively separable (as we did in Chapter 8), one can succeed in finding demand revelation tax schemes: these were nontrivial, nondictatorial and nonmanipulable SCF's.

The third way out of the impossibility theorem is the one that will be dealt with in this and the next chapter. The key will be a change in the way we check for manipulability. Recall that we use R_i to denote person i's preference relation, P_i to denote his strict preference relation, and I_i his indifference relation. Thus far, an SCF F was defined as manipulable if one could find an agent i and preferences $(R_1, \ldots, R_{i-1}, R_{i+1}, \ldots, R_n)$ for the others –we shall use the notation R_{-i} to denote the profile of preferences of all persons but i– such that, when the true preferences of agent i were R_i, he'd rather report R'_i instead, i.e., $F(R'_i, R_{-i}) P_i F(R_i, R_{-i})$. In other words, for nonmanipulability or strategy proofness, we were requiring that, regardless of what the true R_{-i} are, agent i not have an incentive to misrepresent his preferences. This means that reporting the true preferences must be a dominant strategy in the mechanism in which each agent is asked to report his preferences.

Much weaker than requiring truth telling regardless of the other reports is to require truth telling when the others are also telling the truth, and this is what we will utilize in these two chapters. Formally, this is done by appealing to the game theoretic notion of an equilibrium, which was already used in Chapter 8 when the Groves-Ledyard tax scheme was covered. In the current chapter we shall assume complete information among the agents, and the corresponding equilibrium notion is that of Nash equilibrium. In the next chapter this assumption will be relaxed and we shall consider incomplete information environments, in which Bayesian equilibrium will be used.

2. An Example

We begin with an example that should be familiar. We go back to the model of a market economy of Chapter 3. For simplicity, we rule out production considerations. That is, our example will be based on a pure exchange economy with no externalities.

Consider the following two-agent two-goods exchange economy. Suppose the initial endowments of the commodities are $\omega_1 = (3, 9)$ for agent 1, and $\omega_2 = (9, 3)$ for agent 2. This is known by the Central Authority. Everyone knows, including the Central Authority, that agent 1's preferences can be represented by the utility function $u_1(x_{11}, x_{12}) = x_{11}x_{12}$. However, the Central Authority (but not agent 1!) is uncertain about agent 2's preferences. For simplicity, suppose that there are only two possibilities: either agent 2's preferences can be represented by the utility function $u_2(x_{21}, x_{22}) = x_{21}x_{22}$ (with a neutral position towards both goods), or they can be represented by $v_2(x_{21}, x_{22}) = x_{21}^2 x_{22}$ (showing a predisposition for good 1).

Suppose that the SCF that the Central Authority would like to implement is the Walrasian or competitive market equilibrium allocation. We next calculate it.

Suppose the economy is the one described by the utility functions u_1 and u_2. Then, letting p represent the competitive price of good 1, and assuming that the price of good 2 is normalized at 1, the competitive equilibrium is described by the following six equations. The first two give agent 1's optimal choice, the third and fourth agent 2's, and the last two take care of market clearing:

$$\mathrm{MRS}_{u_1} = \frac{x_{12}}{x_{11}} = p;$$
$$px_{11} + x_{12} = 3p + 9;$$
$$\mathrm{MRS}_{u_2} = \frac{x_{22}}{x_{21}} = p;$$
$$px_{21} + x_{22} = 9p + 3;$$
$$x_{11} + x_{21} = 12;$$
$$x_{12} + x_{22} = 12.$$

The reader can check that the solution to this system of equations is $p = 1$, $(x_{11}, x_{12}) = (6, 6)$ and $(x_{21}, x_{22}) = (6, 6)$. Thus, at the competitive equilibrium when agent 2's utility function is u_2, goods are exchanged in a one-to-one ratio and the final allocation is the center of the Edgeworth box.

Let's now calculate the competitive equilibrium when agent 2's utility function is v_2. The corresponding conditions are:

$$\mathrm{MRS}_{u_1} = \frac{x_{12}}{x_{11}} = p;$$
$$px_{11} + x_{12} = 3p + 9;$$
$$\mathrm{MRS}_{v_2} = \frac{2x_{22}}{x_{21}} = p;$$
$$px_{21} + x_{22} = 9p + 3;$$
$$x_{11} + x_{21} = 12;$$
$$x_{12} + x_{22} = 12.$$

The solution is $p = \frac{13}{9}$, $(x_{11}, x_{12}) = (\frac{60}{13}, \frac{20}{3})$, $(x_{21}, x_{22}) = (\frac{96}{13}, \frac{16}{3})$. That is, given agent 2's stronger preference for good 1, it become more valuable and now more units of good 2 have to be paid per unit of good 1 in the market. Agent 2 ends up consuming substantially more good 1 than before, while agent 1 ends up with more good 2.

If the Central Authority knew agent 2's preferences, it would want to implement the allocation $((6, 6), (6, 6))$ when his utility function is u_2, and $((\frac{60}{13}, \frac{20}{3}), (\frac{96}{13}, \frac{16}{3}))$ when it is v_2. Unfortunately, the Central Authority does not have this information. It could try to ask agent 2 what his preferences are, much in the spirit of what we were doing in the last chapter. But then, it would not receive a truthful report. The reason is that $u_2(96/13, 16/3) > u_2(6, 6)$. That is, when his true utility function is u_2, agent 2 has an incentive to report that it is v_2.

In the jargon of the last chapter, the competitive equilibrium SCF is not strategy proof. The question we ask now is whether the Central

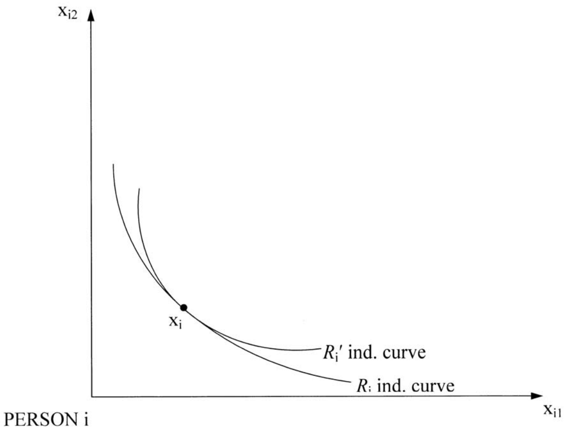

Figure 15.1.

Authority can exploit the fact that agent 1 knows agent 2's true preferences. Perhaps a more sophisticated mechanism can be designed where this bit is used in order to elicit the information truthfully.

As will be shown in the next section, the key to a successful answer to this question is a new requirement on SCF's that was first introduced in Maskin (1999). It is called Maskin monotonicity.

3. Maskin Monotonicity

We begin by defining the central condition in this chapter.

Maskin monotonicity. Suppose that the SCF assigns alternative x when the preference profile is $R = (R_1, R_2, \ldots, R_n)$, i.e., $F(R) = x$. If the preferences of each individual i change from R_i to R_i' in a monotonic way around x (that is, whenever xR_iy, one has that $xR_i'y$), then the alternative socially chosen should not change: $F(R') = x$.

Maskin monotonicity is related to the condition of independence-monotonicity used in the last section of Chapter 13 (we will be specific about how both relate in the appendix to this chapter). The meaning of Maskin monotonicity is illustrated in Figure 15.1. In it, we are representing two indifference curves for an agent in an exchange economy with two goods. In such an example, an "alternative" is simply a "feasible allocation" of goods in the economy. Then, the change in preferences contemplated by the Maskin monotonicity requirement means that the lower contour set at x expands. (Recall from Chapter 1 that the lower contour set of R_i at x is the set $\{y|xR_iy\}$.) To reiterate, suppose that the alternative initially chosen is x. Further, suppose preferences change, but in such a way that for no individual it is true that x has fallen with respect to any other alternative in his personal ranking. Then, Maskin monotonicity says that the social choice should remain x. Of course, if preferences change in any other way (so that at least for one agent the lower contour sets at x are not nested as in Figure 15.1), Maskin monotonicity does not restrict the social choice at all.

Let's now go back to the kind of exchange economy that we were dealing with in our example. Specifically, consider a class of exchange economies with n agents. Assume there are m infinitely divisible commodities; each agent holds positive amounts of each good in his initial endowment and final consumption bundles are represented by m-dimensional vectors with all entries being nonnegative. Suppose preferences are represented by monotonic utility functions whose indifference curves are convex. Suppose also that we have made enough assumptions to guarantee that there is only one competitive equilibrium allocation and that this prescribes positive consumption of all goods for each agent (we will have more to say about this assumption at the end of the chapter).

We now claim that, over the considered class of exchange economies, the competitive equilibrium SCF satisfies Maskin monotonicity. To see this, consider an economy in which $\{1, \dots, n\}$ is the set of agents, agents' preferences are $R = (R_1, \dots, R_n)$ and their initial endowments are $(\omega_1, \dots, \omega_n)$. Denote the competitive equilibrium SCF by F, and let $F(R) = x$, i.e., $x = (x_1, \dots, x_n)$ is the unique competitive equilibrium allocation of this economy, where x_i is the final consumption bundle assigned to agent i in equilibrium. This implies that there exist competitive equilibrium prices p such that the bundle x_i is the optimal choice for each agent over the budget set determined by prices p and endowment ω_i. Furthermore, $\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} \omega_i$.

Next consider an economy consisting of the same agents with the same initial endowments, but in which preferences for agent i have gone through a monotonic change around x. That is, for any bundle $y_i P_i' x_i$, it was already true that $y_i P_i x_i$. Now, because x was a competitive equilibrium allocation in the economy with preferences R, any such bundle lies outside of the budget set for agent i and prices p. Therefore, for each agent $i = 1, \dots, n$ x_i is also his optimal choice over the budget set determined by p and ω_i, even when his preferences are R_i'. Since $\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} \omega_i$, i.e., market clearing still holds, x is a competitive equilibrium allocation in the economy with preferences R': $F(R') = x$. But this means that the competitive equilibrium SCF F satisfies Maskin monotonicity, as we claimed.

4. Maskin's Theorem

The relevance of the requirement of Maskin monotonicity comes from Maskin's theorem, which we will present in this section. But first, we introduce some important notions.

To begin with, let's set up the implementation problem. There is a set of agents $N = \{1, \ldots, n\}$, a set of social alternatives A, and each agent has a preference relation R_i over the set A. Suppose the Central Authority wishes to implement a given SCF F for this society. Suppose the Central Authority knows what is feasible, i.e., it knows the set A. However, it does not know the preferences $R = (R_1, \ldots, R_n)$ of the individuals over the feasible alternatives. Thus, if the true preferences are R, it would like to implement alternative $F(R)$. But the question is: how can this be done without knowing R? Of course, there are certain SCF's for which this is an easy problem. If the SCF is constant, so that $F(R) = x$ for a fixed x no matter what R is, the Central Authority can simply enforce x. But SCF's that we are interested in for societies are not like this: we want our SCF to be sensitive to the preferences of individuals in society.

Next, we need to talk about how the agents will communicate with the Central Authority. In the analysis of the previous chapter, the implicit communication was one in which each agent reported his preferences to the Central Authority. Thus, if agent i's true preferences were R_i, he chose to report $\hat{R}_i$, where this could be a truthful or nontruthful report. With the information collected, i.e., $\hat{R} = (\hat{R}_1, \ldots, \hat{R}_n)$, the Central Authority implemented $F(\hat{R})$. This is a particular kind of communication, but there is no reason to restrict attention to such schemes: in an auction or market context, agents are asked to put a bid on the table instead of reporting their entire preferences, or in voting, agents may be asked simply for their top-ranked alternatives instead of having to report their entire rankings. This leads to the general notion of a mechanism.

A *mechanism* or *game form* G is a pair $G = ((M_i)_{i \in N}, g)$, in which M_i is a set of messages $\hat{m}_i$ that agent i can send to the Central Authority, and g is the outcome function. The function g collects the profile of messages $\hat{m} = (\hat{m}_1, \ldots, \hat{m}_n)$ sent by the agents, and delivers a feasible outcome $g(\hat{m}) \in A$.

A mechanism is also called a *game form*, not a game, because its specification is completely independent of agents's preferences or payoffs. Hence, a game form maps profiles of messages into feasible outcomes. In contrast, a *game* assigns a profile of payoffs or utilities to each profile of messages. Recall that the Central Authority knows the set A, but not the agents' preferences over A. Thus, while the Central Authority can design a game form, it cannot design a game (as the latter requires information that it does not have).

In this chapter, we assume that there is *complete information* among the agents of the set N. That is, each agent i knows his own preferences R_i, and also knows the preferences R_{-i} of the others. Further, this is common knowledge among them: everyone knows everything, everyone knows that everyone knows everything, everyone knows that everyone knows that everyone knows everything, and so on. The complete information assumption is not often met in reality, although it is plausible if one is considering societies with a small number of individuals, who know each other well, much better than an outsider (the Central Authority) may know them. (E.g., think of a small community of neighbors upon whom Town Hall is trying to levy taxes. Or of a union of countries like the European Union from whom the United Nations may want to elicit payments to fund programs in the developing world. Or consider a father who wants to bring about the optimal allocation of toy time sharing among his two daughters, who have a much better idea about their true preferences for toys.) In general, outside of cases like these, we view the complete information assumption as a way to approximate situations in which informational asymmetries among the agents are small, compared to informational asymmetries between the agents on the one hand and the Central Authority on the other hand.

What is interesting now is that the mechanism or game form designed by the Central Authority to govern the communication with the agents is taken by these as a true game. That is, each of them knows his own preferences or payoffs, as well as those of the others. Thus, in choosing the message $\hat{m}_i$ out of the set M_i, strategic considerations in this game of complete information matter. In fact, because of this, we shall use the word *strategy* to talk about a message sent by agent i.

How does agent i choose his strategy $\hat{m}_i$ in the mechanism? Well, of course the answer will depend on his preferences over outcomes, as well as on the specification of the outcome function. Thus, a more appropriate question is: how does he choose his strategy in the game induced by the mechanism G when the preferences are R?

In such a game, he could choose a strategy whenever it is dominant. Strategy $\hat{m}_i$ is dominant whenever $g(\hat{m}_i, m_{-i}) R_i g(m_i, m_{-i})$ for all messages $m_{-i} = (m_1, \ldots, m_{i-1}, m_{i+1}, \ldots, m_n)$ sent by the others and for all $m_i \in M_i$ (with at least one strong preference for each m_i). That is, agent i can never go wrong by choosing $\hat{m}_i$ if it is dominant, because, regardless of what the others choose to do, $\hat{m}_i$ always yields an outcome that is at least as good (and sometimes strictly better) than the outcome produced by any other message sent by agent i. This is a very strong property: often dominant strategies do not exist. In most games, agents will not have it so easy, and their optimal strategy will not be independent of what the others do. Moreover, insisting on dominant strategies in the implementation problem leads us to the Gibbard-Satterthwaite impossibility theorem.

Since the concept of dominant strategies is too demanding, we shall weaken our game theoretic solution concept to that of an equilibrium.

A *Nash equilibrium* of the game induced by the mechanism G and preferences R is a profile of messages $m^* = (m_1^*, \ldots, m_n^*)$ such that for every $i \in N$ $g(m_i^*, m_{-i}^*) R_i g(m_i, m_{-i}^*)$ for every $m_i \in M_i$. That is, at a Nash equilibrium, each agent is choosing a strategy that is optimal given what the others are choosing. This is far weaker than requiring that agent i's strategy be optimal for *any* strategies chosen by the other agents. See our discussion of the concept in the section concerning the Groves-Ledyard tax scheme in Chapter 8.

We shall say that the SCF F is *Nash implementable* whenever one can design a mechanism $G = ((M_i)_{i \in N}, g)$ such that, for every possible preference profile R and for every Nash equilibrium m^* of the game induced by the mechanism G when the preferences are R, $g(m^*) = F(R)$.

Note how Nash implementability does not require that the equilibrium of the game be unique. Indeed, multiple equilibria –strategy profiles– are possible, but all of them must have the same outcome, which should be the socially desirable one. In this chapter our requirement of foolproofness of an SCF amounts to Nash implementability. That is, given that the use of dominant strategies is limited by the Gibbard-Satterthwaite theorem, can one design a mechanism such that, regardless of what the true preferences R are, all Nash equilibria of the corresponding induced game coincide with the one the Central Authority would like to implement if it knew R?

Another way to put the question is this: Can one describe the requirements on SCF's that are equivalent to Nash implementability? The answer is "yes," and the solution was provided by Eric Maskin. We present it in the next two results. The first result will deal with necessity and the second with sufficiency.

> *Maskin's Theorem 1; Necessity*: If the SCF F is Nash implementable, it satisfies Maskin monotonicity.
>
> *Proof*: Since the SCF F is Nash implementable, there exists a mechanism $G = ((M_i)_{i \in N}, g)$ that, when the true preferences are R, has a Nash equilibrium $m^*(R)$ whose outcome is the alternative specified by F: $g(m^*(R)) = F(R) = x$.

Now suppose there is a monotonic change of preferences around x. That is, the new preference profile is R' such that for every $i \in N$, if $yP'_i x$, it was already true that $yP_i x$. Since M^* was a Nash equilibrium of G under preferences R, any unilateral deviation from m^*, such as (m_i, m^*_{-i}) was producing an outcome $g(m_i, m^*_{-i}) = z$ such that $xR_i z$. Therefore, since the mechanism does not vary with a change in preferences, the same outcome z results following the message profile (m_i, m^*_{-i}) when preferences are R'. Because preferences have changed in a monotonic way around x, we know that $xR'_i z$. And since this holds for any unilateral deviation from m^*, this shows that m^* is also a Nash equilibrium of G when preferences are R'. Thus, if preferences are R', we have that x is a Nash equilibrium outcome of the mechanism. But since F is Nash implementable, it must be the case that $F(R') = x$, and then F satisfies Maskin monotonicity. Q.E.D.

Next we state and prove the other direction, almost a converse of the first result. As will become clear in the proof, the mechanism proposed makes use of the Japanese proverb "the nail that sticks up gets hammered down."

Maskin's Theorem 2; Sufficiency: Suppose there are at least three agents and the environment includes a private good. Then, if the SCF F satisfies Maskin monotonicity, it is Nash implementable.

Proof: The proof is based on the construction of a canonical mechanism that will work for any SCF F satisfying Maskin monotonicity, regardless of the implementation problem (exchange economy, production economy, allocation of public goods, voting, etc.)

Consider the following mechanism $G = ((M_i)_{i \in N}, g)$, in which each message $m_i \in M_i$ allowed to agent i consists of an alternative, a preference profile and a nonnegative integer. Thus, a typical message sent by agent i is denoted $m_i = (a^i, R^i, z^i)$. To be clear, $a^i \in A$ is an alternative, $R^i = (R^i_1, \ldots, R^i_i, \ldots, R^i_n)$ is the preference profile (preferences of all agents) reported by i, while $z^i = 0, 1, 2, \ldots$ is a number that i chooses. The outcome function g of the mechanism is defined with the following three rules, where $m = (m_1, \ldots, m_n)$:

(i) If all agents announce the same thing, $m_i = (a^i, R^i, z^i) = (a, R, 0)$ for all $i \in N$, and $F(R) = a$, then $g(m) = a$.

(ii) If there is an almost unanimous announcement as in part (i), i.e., if $n-1$ agents announce $m_i = (a, R, 0)$ with $F(R) = a$, but agent j announces $m_j = (a^j, R^j, z^j) \neq (a, R, 0)$, then we can have two cases:

-If aR_ja^j, then $g(m) = a^j$.
-If a^jP_ja, then $g(m) = a$.

(iii) In all other cases, an integer game is played: identify the agent who announces the highest integer (if there is a tie at the top, pick the one with lowest index among them). This person is declared the winner of the integer game and the alternative implemented is the one that he picks.

We now have to prove two things: (1) $F(R)$ is a Nash equilibrium outcome of this mechanism when the true preferences are R, and (2) there is no other outcome supported by Nash equilibria when the true preferences are R. Therefore, fix an arbitrary preference profile R and let's analyze the game induced by the mechanism G and these preferences.

First, note that the unanimous announcement $m_i = (F(R), R, 0)$ for all $i \in N$ is a Nash equilibrium. These announcements are unanimous in reporting the agents' preferences truthfully, the alternative that is socially desirable under F for these preferences and the integer 0. If these are the announcements, the outcome is decided by rule (i) and it is $F(R)$. Note that unilateral deviations from this announcement cannot induce rule (iii), but only rule (ii). So suppose agent j considers deviating from the unanimously announced message, and instead announces $(a^j, R^j, z^j) \neq (F(R), R, 0)$. The outcome would then be determined by rule (ii). But then, notice that under rule (ii) the outcome would only change to be a^j, the one that j has proposed in his deviation, if aR_ja^j according to preferences R_j. Therefore, since R_j are j's true preferences, agent j will not benefit from such a deviation. It is important to observe how the fact that there are at least three agents is used in this last step, in order to determine what is an "almost unanimous report". This is what allows the mechanism to spot the liar and use the preference R_j for player j, which is being announced by $n-1$ individuals. Note how this would not be possible if one has only two agents: rule (ii) would not be well defined. Can you see why?

Therefore, we have established that the proposed strategy profile is a Nash equilibrium, whose outcome is the "right one," i.e., when the true preferences are R, the outcome is $F(R)$. This proves our goal (1) stated above. The rest of the proof will show that there is no other Nash equilibrium outcome of this mechanism when the true preferences are R, i.e., our goal (2).

To see this, observe first that one cannot have a Nash equilibrium under either rule (ii) or rule (iii). Fix an arbitrary candidate strategy profile that falls under either of these two rules. To prove our claim is simple: if the outcome is determined by either of these two rules, at least $n-1$ individuals are not receiving their top-ranked alternative. The presence of a private good guarantees that the top-ranked alternative for each agent i, call it x^i, is different. But then, any of these individuals can profit from a unilateral deviation: let agent i in this set announce $(x^i, \cdot, z^i)$, where z^i is a large enough integer that is larger than the integers announced by all other agents. With this deviation, agent i will be declared the winner of the integer game and x^i will be implemented, contradicting that the candidate profile was a Nash equilibrium.

Therefore, if there exists another Nash equilibrium of G when the true preferences are R, its outcome must be determined under rule (i). That is, there is a unanimous announcement $m_i = (F(R'), R', 0)$ for all $i \in N$, where $R' \neq R$. That is, the agents are unanimously reporting a false preference profile, but the deception is sophisticated in that the alternative they all mention is the one proposed by F for the reported preferences, and they are also all announcing integer 0. Given these reports, rule (i) would be applied and the outcome would be $F(R')$.

Well, if $F(R') = F(R)$, this is a collective lie that does not bother the Central Authority in the least. After all, the desired alternative is still implemented, so that's fine.

However, suppose that $F(R') \neq F(R)$. Then, this is a collective lie that interferes with the social goals, so this is to be taken seriously. But recall that the SCF F satisfies Maskin monotonicity. Since the alternative chosen by F has changed when preferences are R or R', i.e., $F(R') \neq F(R)$, this implies that alternative $F(R')$ must have fallen in the preference ranking of at least one individual with respect to some other alternative, in going from R' to R. That is, there exists agent j and alternative y such that $F(R')R'_j y$ and $yP_jF(R')$.

Recall that the candidate Nash equilibrium was $m_i = (F(R'), R', 0)$ for all $i \in N$. However, consider the following unilateral deviation announced by agent j: $(y, \text{anything}, \text{anything})$. Note then that the resulting outcome is determined by rule (ii). Moreover, the outcome implemented is y, because $F(R')R'_j y$. But this is great for agent j, whose true preferences are R_j, because $yP_jF(R')$, thereby contradicting the supposition that the candidate profile was a Nash equilibrium.

In conclusion, the only Nash equilibrium outcome of the mechanism G when the true preferences are R is $F(R)$. Since R was any arbitrary preference profile, this shows that the SCF F is Nash implementable and the proof is complete. Q.E.D.

5. Comments on Maskin's Theorem

Maskin's theorem provides an almost complete characterization of the SCF's that are Nash implementable. As such, it is a fundamental result that gives the solution to the implementation problem when there is complete information among the agents and they are assumed to play optimally given what they expect the others to do. That is, agents are assumed to play according to a Nash equilibrium, in which actions and expectations confirm each other. The expectations held by each agent are correct given the equilibrium actions, and given those expectations, the action taken by each agent is optimal.

It follows from the results proved in the previous section that, for implementation problems involving at least three agents and in which there is a private good, Nash implementability of an SCF is equivalent to Maskin monotonicity. This result allows implementation theory to move beyond the impossibility theorem of Gibbard and Satterthwaite. For example, as already argued earlier in this chapter, there are interesting classes of economies in which the competitive SCF will be Maskin monotonic, and therefore, in those domains it is Nash implementable.

As pointed out in the proof of the sufficiency part of the theorem, the requirement of there being at least three agents is used in the canonical mechanism. Indeed, the case of Nash implementation for two agents requires an additional condition. The reason is simple: it is easier to catch a liar in a community of at least three agents (by pointing fingers at the liar) than in one of only two ("my word against yours").

In cases where there is no private good, it is not possible to bribe people, for example, by offering more money. Then, to preserve Maskin's Theorem 2, the requirement of *weak no veto* on the SCF must be added. This requirement says that if at least $n - 1$ individuals agree that an alternative is top-ranked, the SCF should choose it. It turns out that, while weak no veto is not necessary for Nash implementability, it is sufficient, together with Maskin monotonicity, for problems where there is no private good.

Finally, we are presenting this material for single-valued SCF's, for which a single alternative is picked out for each preference profile. The theory can be extended to multi-valued SCF's. A multi-valued SCF is called a *social choice correspondence.* Maskin's theorem continues to apply. For it, the condition of Maskin monotonicity must be adapted as follows: if x is one of the alternatives chosen by the SCF F for preferences R and there is a monotonic change of preferences around x from R to R', then x must continue to be one of the alternatives chosen by F at profile R'. Also, the definition of Nash implementability now means that one can find a mechanism such that, for each preference profile, the set of its Nash equilibrium outcomes coincides with the set selected by the SCF. With essentially the same proof as in the previous section, one can show that for problems involving at least three agents and a private good, a multi-valued SCF is Nash implementable if and only if it satisfies Maskin monotonicity. It turns out that many correspondences of interest satisfy Maskin monotonicity, including the *Pareto optimality correspondence*, the *core* and a restriction of the Walrasian equilibrium correspondence called the *constrained Walrasian correspondence* (see the exercises section).

To illustrate the use of Maskin's theorem, consider the exchange economy example that we saw earlier in this chapter, but now let's add a third agent. Recall that agent 1's utility function is $u_1(x_{11}, x_{12}) = x_{11}x_{12}$ and his endowment is $\omega_1 = (3, 9)$. Agent 2's endowment is $\omega_2 = (9, 3)$, but his utility function could be either $u_2(x_{21}, x_{22}) = x_{21}x_{22}$ or $v_2(x_{21}, x_{22}) = x_{21}^2 x_{22}$. The new agent, agent 3, has a utility function $u_3(x_{31}, x_{32}) = \min\{x_{31}, x_{32}\}$ and his endowment is $\omega_3 = (8, 8)$.

You can check as an exercise that the only Walrasian equilibrium allocation in each of the two possible economies is the same as before: as a function of agent 2's utility function, the equilibrium bundles for agents 1 and 2 are the ones given before, while agent 3 receives his endowment bundle. You can also check that the only Nash equilibrium of the canonical mechanism in each of the two possible economies has each agent reporting the true utility function of agent 2's, the Walrasian equilibrium allocation for the true economy, and the integer 0. This implements the Walrasian bundle for each agent in each of the two economies.

Therefore, the Central Authority, by making the three agents send messages using this mechanism, will be able to bring about the Walrasian equilibrium allocation in each economy.

6. Limitations of Maskin Monotonicity and Approximate Implementation

We have argued that some interesting SCF's are Maskin monotonic, but it is true that some are not. Therefore, given Maskin's Theorem 1, there will be limits to the success of Nash implementability. For instance, consider the following example, taken from the First Book of Kings in the Bible.

The wise King Solomon was presented with the following implementation problem (1st Kings, chapter 3, verses 16-28). Two women, whom we shall call A and B, each claim to be the true mother of one baby. King Solomon, the Central Authority, wants to implement the SCF that allocates the baby to its true mother. However, he does not know who the true mother is.

Let's model the problem as follows. Following the Bible story, suppose that there are three possible alternatives: a (the baby is allocated to woman A), b (the baby is allocated to B) and c (the baby is divided with a sword, with half given to A and the other half to B).

Of course, the true mother does not want to see her baby cut in half. The true mother views c as the worst outcome, but the false mother does not. We assume that the preference profile corresponding to A being the true mother is $R = (R_A, R_B)$, given by the following table (all preferences are strict):

P_A	P_B
a	b
b	c
c	a.

Similarly, preference profile $R' = (R'_A, R'_B)$ corresponds to B being the true mother, given by the table below (again all preferences are strict):

P'_A	P'_B
a	b
c	a
b	c.

In our terminology, the SCF F that King Solomon wants to implement is the following: $F(R) = a$ and $F(R') = b$. But now we claim that F violates Maskin monotonicity. Indeed, since $a = F(R) \neq F(R') = b$, for F to satisfy Maskin monotonicity, it would be necessary that in the preference change from R to R' there is an agent i and an alternative that has risen in R'_i with respect to $a = F(R)$. But this agent is nowhere to be found: for agent A, a is top ranked in both preference profiles, and for B alternative a actually rises with respect to c in going from R_B to R'_B. In other words, the preference change from R to R' is a monotonic change around a, and therefore, Maskin monotonicity would require the social choice to stay put at a, but this does not happen.

It follows from the necessity part of Maskin's theorem that the Solomonic SCF F is not Nash implementable. That is, there does not exist any mechanism that, making use of the complete information existing between the two women, has as its only Nash equilibrium outcome the one recommended by the SCF F. In short, implementing the Solomonic SCF is not trivial. It is impossible to do in Nash equilibrium.

Fortunately for Solomon, however, the false mother acted foolishly. She announced her true preferences, with c in the middle of her ranking instead of at the bottom: "But the other [woman] said: 'Let it be neither mine nor thine, but divide it.' " So, she gave herself away as the false claimant, and then Solomon gave the baby to the true mother.

If the false mother had put c at the bottom of her ranking, like the true mother, Solomon would have failed as Central Authority. If both women were fully strategic in playing the mechanism, Solomon could not have accomplished his goal of allocating the baby to its true mother since this SCF fails Maskin monotonicity.

One way out of the limitations imposed by Maskin monotonicity is the approach of *approximate implementation.* To talk about approximate implementation, we shall introduce lotteries over alternatives. That is, if the set of alternatives is $A = (a_1, \ldots, a_k)$, we let $(q_1, \ldots, q_k)$ be a lottery over alternatives. For $j = 1, \ldots, k$, alternative a_j is implemented with probability q_j. Of course, $q_j \geq 0$ for $j = 1, \ldots, k$ and $\sum_{j=1}^{k} q_j = 1$. The interpretation is that the Central Authority may now use a random device by which each alternative is implemented with some probability.

Instead of *exact implementation*, as we had so far, in which for each preference profile R alternative $F(R) \in A$ was implemented with probability 1, we shall now speak of *approximate implementation*: for any arbitrarily small $\epsilon > 0$ and for any preference profile R, alternative $F(R)$ is implemented with probability $1 - \epsilon$.

To evaluate lotteries, we shall assume that agents have preferences that can be represented by *expected utility* functions. The reader is referred to the relevant section of Chapter 1. In the present context, for each agent i there exist numbers $u_i(a_j)$ for each $j = 1, \ldots, k$ (agent i's utilities associated with each pure alternative), such that the utility that agent i derives from a lottery $l = (q_1, \ldots, q_k)$ is its expected utility, i.e., $u_i(l) = \sum_{j=1}^{k} q_j u_i(a_j)$. Now recall that the indifference curves corresponding to expected utility are parallel straight lines. As an illustration, Figure 15.2 depicts the probability simplex for the case of three alternatives (a_1, a_2, a_3) and corresponding utilities $u_i(a_1) = 2$, $u_i(a_2) = 1$ and $u_i(a_3) = 0$ according to preferences R_i. For these preferences, the indifference curve of level $\bar{u}$ is the locus of points in the simplex whose equation is $2q_1 + q_2 = \bar{u}$. Not surprisingly, the top ranked point in the simplex is the degenerate lottery that puts all the weight on a_1, while the worst lottery is the degenerate one with all the weight on a_3.

Figure 15.2 also shows an indifference map with different expected utility preferences over lotteries. In it, $u_i'(a_1) = 4$, $u_i'(a_2) = 3$ and

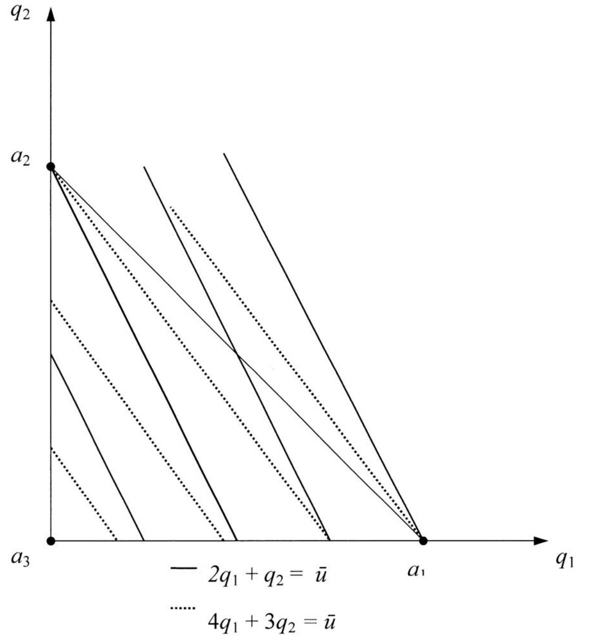

Figure 15.2.

$u_i'(a_3) = 0$, and we call these preferences R_i'. For them, the indifference curve of level $\bar{u}$ has the equation $4q_1 + 3q_2 = \bar{u}$.

Recall that, despite the fact that u_i' is a monotone transformation of u_i, both utility functions do not represent the same preferences over lotteries.

Now consider SCF's whose range is the interior of the probability simplex, i.e., SCF's that assign to each preference profile R a lottery $F(R)$ that puts positive probability on each pure alternative. Therefore, the point $F(R)$ cannot be on the sides of the probability simplex, but must be contained in its interior. Now it is easy to see that any such SCF satisfies Maskin monotonicity. To see this, it suffices to observe that no preference change is a monotonic change around such $F(R)$: when preferences over lotteries change from R to R', it is never the case that the lower contour set of lotteries at an interior point $F(R)$ when preferences are R is contained in the lower contour set of the same $F(R)$ when preferences are R'. Therefore, Maskin monotonicity does not impose any restriction on what F should be for preferences R'.

But then, using Maskin's theorem, we have established that if there are at least three agents and a private good, every SCF whose range is in the interior of the probability simplex is (exactly) Nash implementable. This implies that every SCF (even those whose range includes the sides of the simplex) is *approximately* Nash implementable. This is a very insightful result, first discovered by Dilip Abreu, Arunava Sen and Hitoshi Matsushima in two papers written independently in the late 1980's. If we allow approximate implementation, we obtain a universal possibility result.

7. Exercises

1. In exchange economies where the initial endowment is $(\omega_i)_{i \in N}$, consider the following SCF F. Let x^* be a feasible allocation. Then, let $F(R) = x^*$ if x^* is Pareto optimal when the preferences are R, and $F(R) = (\omega_1, \ldots, \omega_n)$ otherwise. Is this SCF Maskin monotonic? Provide a proof for your answer.

2. Recall the Groves-Ledyard mechanism of Chapter 8. This was proposed in order to implement the allocation that satisfies the necessary Samuelson condition of efficiency, and it was a self-funded mechanism. Show that the SCF consisting of the described allocation rule satisfies Maskin monotonicity.

3. Consider King Solomon's problem. One alternative elaboration of the story is the following (we say it is an elaboration because the Bible does not describe the outcome that Solomon would have implemented for every possible contingency that could have happened). Suppose that Solomon thought first of using mechanism $G_1 = ((M_i)_{i=A,B}, g_1)$, where $M_i = \{\hat{A}, \hat{B}\}$ is simply a declaration of who is the true mother, and the outcome function g_1 was:

$$g_1(\hat{A}, \hat{A}) = a;$$
$$g_1(\hat{B}, \hat{B}) = b;$$
$$g_1(\hat{A}, \hat{B}) = g_1(\hat{B}, \hat{A}) = c.$$

   a Illustrate why this mechanism does not work to Nash implement the Solomonic SCF. Specifically, find the Nash equilibrium outcomes of the game induced by the mechanism when A is the true mother and when B is the true mother.

   b However, when he asked the women and the reports turned out to be $(\hat{A}, \hat{B})$, he was ready to implement outcome c by turning to his sword. Then, the true mother (say, A) changed her report to $\hat{B}$. Now, with both reports being $\hat{B}$, Solomon did not use the outcome function g_1. Instead, he implemented $g_2(\hat{B}, \hat{B}) = a$.

The Bible does not tell us the rest of the outcome function g_2, but perhaps woman B could appeal his wise decision: "eh! your Wise Majesty is not using g_1!" If this fact had been known before the women sent their messages, a different strategic analysis would have probably led to a different report profile. The problem of course is that we are not given the complete description of the mechanism that King Solomon was using.

So, complete the mechanism based on outcome function g_2 as follows:

$g_2(\hat{A}, \hat{A}) = g_2(\hat{B}, \hat{B}) = a;$
$g(\hat{A}, \hat{B}) = g_2(\hat{B}, \hat{A}) = c.$

Evaluate this mechanism from the point of view of implementation theory if one wishes to use it in order to implement the Solomonic SCF.

4. Show that the weak Pareto correspondence of any implementation problem satisfies Maskin monotonicity. (The weak Pareto correspondence when preferences are R prescribes the set of all its weak Pareto optimal alternatives, i.e., the set of feasible alternatives x such that there does not exist another feasible alternative y that every agent *strictly* prefers to x).

5. Show that the Pareto correspondence of any implementation problem need not satisfy Maskin monotonicity. (The Pareto correspondence when preferences are R prescribes the set of all its Pareto optimal alternatives as defined in Chapter 2, i.e., the set of feasible alternatives x such that there does not exist another feasible alternative y that every agent *weakly* prefers, and at least one agent *strictly* prefers, to x). Show, however, that in exchange economies with continuous and monotone preferences the Pareto correspondence satisfies Maskin monotonicity.

6. Recall the core, a solution concept introduced in Chapter 2, as the set of coalitionally stable allocations of an exchange economy. Define its weak version based on strict blocking (as we have just done in exercise 4 for the Pareto correspondence). Show that the weak core correspondence satisfies Maskin monotonicity.

7. Show that the correspondence that assigns to each exchange economy the set of all its Walrasian or competitive equilibrium allocations may violate Maskin monotonicity. On the other hand, define the constrained Walrasian equilibrium correspondence as follows. Let A_i denote the set of bundles x_i for agent i such that for each good j,

$0 \leq x_{ij} \leq \sum_{i=1}^{n} \omega_{ij}$. Given prices p and agent i's endowment ω_i, his budget set $B_i(p)$ is the set of bundles $x_i \geq 0$ such that $p \cdot x_i \leq p \cdot \omega_i$. An allocation x is a constrained Walrasian equilibrium allocation if $\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} \omega_i$ and for each agent i x_i maximizes i's utility over the set $B_i(p) \cap A_i$. Show that the constrained Walrasian equilibrium correspondence satisfies Maskin monotonicity.

8. Appendix

This appendix establishes connections between the properties of independence-monotonicity and Maskin monotonicity. A version of independence-monotonicity was used in Chapter 13 to prove a general version of Arrow's impossibility theorem.

Notational preliminaries: $\bar{A}$ is the set of all alternatives. Alternatives are x, y, z, etc. $\mathcal{A}$ is the set of subsets of $\bar{A}$. A, B, C, etc. are subsets of the set of all alternatives. R_i is person i's preference relation; P_i indicates strict preference for i. $R = (R_1, R_2, \ldots, R_n)$ is a preference profile. R_s is a social preference relation; P_s indicates strict preference for society. We assume that all the R_i's are complete and transitive. A social preference relation R_s may or may not be complete and transitive.

Definitions: An *Arrow social welfare function* is a mapping from the set of preference profiles into the set of social preference relations.

An Arrow social welfare function satisfies $I-M$ (independence-monotonicity) if the following holds:

For any pair of alternatives $\{x, y\}$ and any pair of profiles R and R', if $xP_iy \Rightarrow xP'_iy$ for all i and $yP'_ix \Rightarrow yP_ix$ for all i, then $xP_sy \Rightarrow xP'_sy$.

A *social choice function* is a mapping from the set of preference profiles into the set of all alternatives $\bar{A}$. A *generalized social choice function* is a mapping from the set of preference profiles $\times \mathcal{A}$ into $\bar{A}$; in particular, it takes a preference profile R and a subset A, and produces one element of A.

We write $F(R)$ for a social choice function, and $F(R, A)$ for a generalized social choice function.

We say a social choice function $F(R)$ satisfies *Maskin monotonicity* if for any pair of alternatives $\{x, y\}$ and any pair of profiles R and R', if $xR_iy \Rightarrow xR'_iy$ for all i, then $F(R) = x \Rightarrow F(R') = x$. Given that individual preference relations are assumed complete, $xR_iy \Rightarrow xR'_iy$ is equivalent to $yP'_ix \Rightarrow yP_ix$. Therefore $F(R)$ satisfies Maskin monotonicity if for any pair of alternatives $\{x, y\}$ and any pair of profiles R and R', if $yP'_ix \Rightarrow yP_ix$ for all i, then $F(R) = x \Rightarrow F(R') = x$.

We say that a generalized social choice function $F(R, A)$ satisfies *Maskin monotonicity* if for any pair of alternatives $\{x, y\} \subset A$, and any pair of profiles R and R', if $yP'_ix \Rightarrow yP_ix$ for all i, then $F(R, A) = x \Rightarrow F(R', A) = x$.

A generalized social choice function $F(R, A)$ can be used to define a strict social preference relation P_s as follows: Say xP_sy whenever $F(R, \{x, y\}) = x$. We'll call P_s the social preference relation *induced* by $F(R, A)$. Such a social preference relation can be generated for any preference profile R. The rule that transforms preference profiles into social preference relations in this fashion will be called the Arrow social welfare function *induced* by $F(R, A)$.

> *Proposition 1.* Suppose the generalized social choice function $F(R, A)$ satisfies Maskin monotonicity. Then the Arrow social welfare function induced by F satisfies $I - M$.
>
> *Proof*: Suppose for any pair of alternatives $\{x, y\}$, and any pair of profiles R and R', the following holds:
>
> $$\text{(a) } xP_iy \Rightarrow xP'_iy \text{ for all } i \text{ and}$$
> $$\text{(b) } yP'_ix \Rightarrow yP_ix \text{ for all} i.$$
>
> We need to show $xP_sy \Rightarrow xP'_sy$ for the induced strict social preference relations P_s and P'_s. Let xP_sy, i.e., $F(R, \{x, y\}) = x$. We want to show that xP'_sy, i.e., $F(R', \{x, y\}) = x$.
>
> At the preference profile R' restricted over the pair $\{x, y\}$, for any individual i, either xR'_iy or yP'_ix. But in the latter case, yP_ix, by (b) above. Thus, restricted over the pair $\{x, y\}$, the change in preferences from R to R' has been a monotonic change around x: alternative y has not become strictly better than x for any agent i at the profile R'. Thus, since the generalized social choice function satisfies Maskin monotonicity, $F(R', \{x, y\}) = x$. Therefore, the Arrow social welfare function induced by F satisfies I-M. Q.E.D.

We now turn to a near converse proposition.

We will now restrict our attention to Arrow social welfare functions that map into the set of *complete*, *transitive* and *strict* social preference relations. Let P_s be the social preference relation produced by such an Arrow social welfare function. Then, $\bar{A}$ has a unique top-ranked alternative under P_s, that is, there is an x such that xP_sy for all $y \neq x$ in $\bar{A}$.

We define the social choice function $F(R)$ *induced* by the Arrow social welfare function by $F(R) = x$, where x is the top-ranked alternative under P_s.

In the proposition that follows, we also assume $\bar{A}$ is the set of allocations in an exchange economy, and we make the assumptions that we had made in Chapter 3 regarding individuals' preferences. That is, we assume that all the individuals have continuous, monotonic, and self-interested preferences.

> *Proposition 2.* Let $\bar{A}$ be the set of allocations in an exchange economy. Suppose an Arrow social welfare function maps into the set of complete, transitive and strict social preference relations. Let $F(R)$ be the induced social choice function. Assume the Arrow social welfare function satisfies I-M.
>
> Then, $F(R)$ satisfies Maskin monotonicity.
>
> *Proof.* Suppose that $F(R) = x$, which means that xP_sy for all $y \neq x$. Now suppose that there is a monotonic change of preferences around x. That is, we consider a preference profile R' such that for all agents i and all y, $yP'_ix \Rightarrow yP_ix$.
>
> We argue by contradiction. Suppose that F does not satisfy Maskin monotonicity. This means that $F(R') = z \neq x$. By the definition of F, this implies that zP'_sx. And of course, it was the case that xP_sz.
>
> Because the Arrow social welfare function satisfies I-M, one of the two premises in the definition of I-M (i.e., $xP_iz \Rightarrow xP'_iz$ for all i, and $zP'_ix \Rightarrow zP_ix$ for all i) must not hold. Therefore, either
>
> [a] xP_iz and zR'_ix for some individual i, or
>
> [b] zP'_ix and xR_iz for some i.
>
> However, case [b] is impossible because the preference change from R to R' has been a monotonic change around x. So the only possibility is case [a]. But in exchange economies with continuous, monotonic and self-interested preferences, this case is also impossible, because it would imply that we can find another alternative $\bar{z}$ arbitrarily close to z such that $xP_i\bar{z}$ and $\bar{z}P'_ix$, which would also contradict our assumption that the preference change has been monotonic.
>
> Thus, both cases are impossible and the proof is complete. Q.E.D.

9. Selected References

(Items marked with an asterisk (*) are mathematically difficult.)

*1. D. Abreu and A. Sen, "Virtual Implementation in Nash Equilibrium," *Econometrica* V. 59, 1991, pp. 997-1021.

This is one of the two original articles that introduce the idea of approximate or virtual implementation (see also reference 5).

*2. L. Hurwicz, E. Maskin and A. Postlewaite, "Feasible Nash Implementation of Social Choice Rules when the Designer Does not Know Endowments or Production Sets," in Ledyard, J. O. (ed.) *The Economics of Informational Decentralization: Complexity, Efficiency and Stability*, Kluwer Academic Publishers, Amsterdam, 1995.

This paper studies implementation in contexts in which the Central Authority does not necessarily know the set A of feasible alternatives. The paper contains many examples and the most relevant results concerning the Walrasian and constrained Walrasian correspondences.

*3. M. Jackson, "A Crash Course in Implementation Theory," *Social Choice and Welfare* V. 18, 2001, pp. 655-708.

This is a clear survey on implementation theory.

*4. E. Maskin, "Nash Equilibrium and Welfare Optimality," *Review of Economic Studies* V. 66, 1999, pp. 23-38.

This is Maskin's classic paper, which contains his theorem. The first version of this important article circulated as an MIT working paper in 1977. Evidently, Maskin had some difficulties implementing its publication. Publishing in economics sometimes is associated with interesting funny stories, like this.

*5. H. Matsushima, "A New Approach to the Implementation Problem," *Journal of Economic Theory* V. 45, 1988, pp. 128-144.

This is the other original paper on approximate implementation (see also reference 1).

*6. J. Moore, "Implementation, Contracts and Renegotiation in Environments with Complete Information," in J. J. Laffont (ed.) *Advances in Economic Theory, VI World Congress of the Econometric Society* (vol. I), Cambridge University Press, 1992.

This is an excellent survey on implementation theory under complete information. It is divided into two parts, and the first is less technical. Its coverage of the King Solomon problem is delightful.

*7. R. Repullo, "A Simple Proof of Maskin Theorem on Nash Implementation," *Social Choice and Welfare* V. 4, 1987, pp. 39-41.

This paper contains the proof of the sufficiency part of Maskin's theorem that we have presented.

*8. R. Serrano, "The Theory of Implementation of Social Choice Rules," *SIAM Review* V. 46, 2004, pp. 377-414.

In this survey, one can find a specific mechanism that approximately implements the Solomonic SCF in Nash equilibrium. This can be done, even though the problem involves only two agents.