Lecture 6. Contracting with Unverifiable information

We now go back to the first lecture on mechanism design, and focus on a related problem: the implementation problem. "The implementation problem is the problem of designing a mechanism (game form) such that the equilibrium outcomes satisfy a criterion of social optimality embodied in a social choice rule. If a mechanism has the property that, in each possible state of the world, the set of equilibrium outcomes equals the set of optimal outcomes identified by the social choice rule, then the social choice rule is said to be implemented by this mechanism."[1] A social choice rule specifies, for each possible state of the world, which outcomes would be socially optimal in that state. It can be thought of as embodying the welfare judgements of a social planner.

Implementation theory encompasses very different problems: provision of public or private goods, auction models, monopoly pricing, optimal taxation. Very often the agents' preferences are privately known (as in our first lectures). But there are also many situations in which information is complete, but nevertheless the implementation problem is not trivial. This is the case whenever information is observable but not verifiable, that is, it is not possible to write binding contracts that have clauses that depend on the state of the world.

Moore (1992) provides some examples. Consider a club whose members want to write the club's rules. These rules prescribe how in the future the members of the club will decide on the use of the club's budget at the beginning of each year. Say at the beginning of each year a state of the world θ realizes (this could be the conditions of the facilities for a sport club). Let's say that the club wants to implement $f(\theta)$. θ is observable but not verifiable. E.g. a sport club wants to repair each winter whichever facility is the most run down. Even though club members anticipate that next January they will all see whether the tennis court or the football pitch is in worse shape, this is something that is hard to verify objectively, say by a civil judge in case the club members go into litigation. So what club members do is to set up a rule on how to

[1] Maskin and Sjorstrom, 2002, "Implementation Theory. Handbook of Social Choice and Welfare," vol 1.

decide at the beginning of each year: e.g. everyone has to vote (with a monetary punishment in case of no vote).

The club members could avoid specifying any mechanism, and instead negotiate an outcome each January, once θ is known. "However, although this may be ex post efficient, in general there will be inefficiences ex ante. Consider two firms that are not vertically integrated. Let θ_1 index the upstream firm's ex post situation (which may include such things as input prices, productivity, and technology), and let θ_2 index the downstream firm's circumstances (e.g., its productivity, technology, and the state of demand for the final product). As in the club example, we suppose that, although both firms observe $\theta = \{\theta_1, \theta_2\}$ this is not observable to outsiders; so a contract cannot be conditioned on θ. $f(\theta)$ denotes the desired price and quantity of the goods traded, together with their quality, time of delivery, and so forth. An optimal $f(\theta)$ will involve not just efficient ex post trading, but also take into account certain key ex ante efficiency considerations such as risk sharing and/or inducing the firms to make appropriate (non-contractible) specific investments after they have signed the contract ... the point is that the firms must contractually specify a mechanism g which implements $f(\theta)$; leaving decisions to ex post bargaining would in general not lead to the desired outcomes." (Moore, 1992)

Implementability properties hinge crucially on the equilibrium concept.

Nash Implementation

The environment is $\{\Theta, Y, I\}$, Θ is the set of possible states of the world; Y is the set of possible outcomes; I is the set of agents. Each agent has rational preferences over outcomes. Agents' preferences can depend on the state. Information is symmetric: all agents observe the state of the world.

The problem is to design a mechanism such that for every state of the world θ, ***for all Nash equilibria***, the outcome belongs to $f(\theta) \subseteq Y$.[2]

Definition. *$f(\cdot)$ is monotonic if for each pair $\theta \in \Theta$ and $\tilde{\theta} \in \Theta$ and for each $y \in f(\theta)$ we have $y \in f(\tilde{\theta})$ whenever for each player i and for each outcome $z \in Y$, if y is weakly preferred to z*

[2] $f(\theta)$ can be a single element for each θ or include multiple elements for some/all θ - that is, $f(\cdot)$ can be a function or a correspondence.

by player i in state θ (denote this as $y \succsim_i^\theta z$ for all $i \in N$ and all $z \in Y$) then y is also weakly preferred to z by player i in state $\tilde{\theta}$ ($y \succsim_i^{\tilde{\theta}} z$ for all $i \in I$).

Definition. *$f(\cdot)$ satisfies weak no-veto power (WNVP) if, for each state θ, outcome y must be in the acceptable set $f(\theta)$ whenever at most one agent does not have outcome y as her preferred outcome.*

Each mechanism defines a message space M_i for each agent and an outcome function $g : \times_{i=1}^{I} M_i \to Y$.

Definition. *$f(\cdot)$ is Nash implementable if there exists a mechanism for which, in every state θ, **all** (Nash) equilibrium outcomes belong to $f(\theta)$.*

Theorem. *If $f(\cdot)$ is Nash implementable, then it is monotonic. With $|I| \geq 3$ if $f(\cdot)$ is monotonic and satisifes WNVP, then it is Nash implementable.*

Proof.

Necessary condition. The first half of the theorem is proved by contradiction. Suppose $f(\cdot)$ is not monotonic and is implementable. Then there exists a pair of states θ, $\tilde{\theta}$ and an outcome y such that:

- $y \in f(\theta)$, $y \notin f(\tilde{\theta})$,
- for all $i \in I$ and all $z \in Y$, if $y \succeq_i^\theta z$ then $y \succeq_i^{\tilde{\theta}} z$.

As $f(\cdot)$ is assumed to be implementable and $y \in f(\theta)$, then there exists a mechanism M such that y is in the set of Nash equilibrium outcomes for state θ, with some set of equilibrium messages m^*. This implies that for all agents i and all $m_i \in M_i$: $y \succeq_i^\theta g(m_i, m_{-i}^*)$. But then, by assumption, $y \succeq_i^{\tilde{\theta}} g(m_i, m_{-i}^*)$, thus y is a Nash equilibrium outcome in state $\tilde{\theta}$, thus contradicting $y \notin f(\tilde{\theta})$.

Sufficient condition. The second half of the theorem is proved by construction. Consider a mechanism M where the set of messages is $\{(\theta, y, n) | \theta \in \Theta, y \in Y, n \in N\}$ where N denotes the set of all integers, with these properties:

- if all agents announce the same y and θ and $y \in f(\theta)$, then $g(m) = y$;

- if all agents except i announce state θ and outcome $y \in f(\theta)$, while agent i announces state $\tilde{\theta}$ and outcome z, then $g(m) = z$ if $y \succeq_i^\theta z$, and $g(m) = y$ if $z \succ_i^\theta y$;
- in all other cases the outcome of the mechanism is $\tilde{y}$ where outcome $\tilde{y}$ is the one announced by the agent who has announced the highest positive integer n.

You can check that for each state θ and outcome $y \in f(\theta)$, y is the outcome of a Nash equilibrium. All is left is to show that for each state θ if $\tilde{y} \notin f(\theta)$ then $\tilde{y}$ cannot be the outcome of a Nash equilibrium. This can be shown in 3 steps. Suppose state is θ.

1) Assume everybody announces state $\tilde{\theta}$ and $\tilde{y}$ such that $\tilde{y} \in f(\tilde{\theta})$ but $\tilde{y} \notin f(\theta)$. Monotonicity ensures $\exists i \in I$ and a $z \in Y$ such that $\tilde{y} \succeq_i^{\tilde{\theta}} z$ but $z \succ_i^\theta \tilde{y}$. But then by definition of the mechanism agent i can deviate and ensure outcome z thus $\tilde{y}$ is not the outcome of a Nash equilibrium;

2) If $\tilde{y}$ is the result of non-unanimous announcements about state and outcome, then by definition of the mechanism there is at most 1 agent that cannot deviate and impose her preferred outcome. All other agents can do so, by announcing their preferred outcome and an integer higher than the integers all other agents annouce in equilibrium. Thus WNVP ensures that if $\tilde{y}$ is the outcome of a Nash equilibrium then $\tilde{y} \in f(y)$;

3) If $\tilde{y}$ is the result of unanimous announcements about state $\tilde{\theta}$ and outcome $\tilde{y} \notin f(\tilde{\theta})$, then some agent must find it optimal to deviate, announce the largest integer and get her preferred outcome.

QED

Subgame-Perfect Implementation

The problem with Nash equilibria is that often there are too many of those. A way to make the implementation problem easier is to consider a subset of Nash equilibria: Subgame-Perfect Nash equilibria. Adding the requirement of subgame perfection allows to introduce credible threats, thus reducing the number of equilibria.

Here we look at a somewhat special example. Consider a public-good choice problem, with 2 agents $i = 1, 2$ and decision $y \in \{0, 1\}$ ($y = 1$ the good is produced, $y = 0$, it is not). A decision rule also specifies transfers $(t_1, t_2) \in \mathbb{R}^2$. Each agent $i's$ preferences depend on a parameter θ_i, and the state of nature is $\theta = \{\theta_1, \theta_2\} \in \Theta$. Utility of agent i is: $\theta_i y + t_i$. θ is observed by both agents. A decision rule is a triple $\{y(\theta), t_1(\theta), t_2(\theta)\}$.

Under subgame-perfect implementation, monotonicity is not required: any outcome rule can be implemented as a unique subgame-perfect equilibrium, with a mechanism that balances the budget on path and may balance the budget or generate surplus off-path. [3]

Consider a mechanism summarized in the following two stages:

1) agent 1 makes a report about its type: $\hat{\theta}_1^1$; agent 2 observes the report and makes a report about the type of agent 1: $\hat{\theta}_1^2$; if $\hat{\theta}_1^1 = \hat{\theta}_1^2$ the game proceeds to step 2; if instead if $\hat{\theta}_1^1 \neq \hat{\theta}_1^2$, then agent 1 can choose between two options: either $\{x, t_x - \Delta t, -t_x - \Delta t\}$ or $\{z, t_z - \Delta t, -t_z + \Delta t\}$, where Δt is large and:

$$\begin{aligned} \hat{\theta}_1^1 x + t_x &> \hat{\theta}_1^1 z + t_z, \\ \hat{\theta}_1^2 x + t_x &< \hat{\theta}_1^2 z + t_z. \end{aligned}$$

2) (if $\hat{\theta}_1^1 = \hat{\theta}_1^2$): same as step 1, with reversed roles. If $\hat{\theta}_2^2 = \hat{\theta}_2^1$ then outcome $\left\{y\left(\hat{\theta}_1^1, \hat{\theta}_2^2\right), t_1\left(\hat{\theta}_1^1, \hat{\theta}_2^2\right), t_2\left(\hat{\theta}_1^1, \hat{\theta}_2^2\right)\right\}$ is implemented.

One way to satisfy these properties is to choose $x = 1 - z$, and set $x = 1$ if $\hat{\theta}_1^1 > \hat{\theta}_1^2$, and $x = 0$ if $\hat{\theta}_1^1 < \hat{\theta}_1^2$.

[3] If there where more than 2 agents, it would be possible to implement with a mechanism that balances the budget both on and off path.