Handout 5
Econ 533
February 26, 2016
TA: Blake Riley

1 Mechanism Design

With social choice theory, we looked at how to aggregate preferences assuming these were observable. Gibbard-Satterthwaite points us toward the direction of imperfectly observable preferences that must be revealed by the individuals before they can be aggregated.

We again consider a set of agents n which will make a collective choice form a set of alternatives X. Utility functions $u_i(x,\theta_i)$ are commonly known, but are determined by a pivately known random variable $\theta_i \in \Theta_i$ which is the agent's type. Typically there is a commonly known prior $\phi(\cdot)$ representing a density over $\theta \in \Theta_1 \times \ldots \times \Theta_n$. Since the utility functions themselves are common knowledge in this context, all uncertainty handled is through the type. Types can also encode the beliefs of an agent. Our new formulation of a social choice function is a straightforward extension based on this.

Definition 1.1: *A **social choice function** is a rule $f : \Theta_1 \times \ldots \times \Theta_n \to X$ that assigns a chosen element $f(\theta) \in X$ for each profile of individual preferences derived from the vector of types θ.*

Some properties are also simple extensions:

Definition 1.2: *The social choice function $f : \Theta \to X$ is **ex-post efficient** if for all $\theta \in \Theta$ there does not exist an $x \in X$ such that $u_i(x,\theta_i) \geq u_i(f(\theta),\theta_i)$ for all i and $u_i(x,\theta_i) > u(f(\theta),\theta_i)$ for some i.*

2 Mechanisms and Implementation

Since agents' types are private information, they must communicate this information to the rest of the group or a central body. We could have any theory about how agents choose the messages they will send, but we typically assume messages will be chosen strategically. Once the messages are collected, an outcome will be chosen by the group. Together, the possible messages and the mapping from messages to outcomes define a mechanism.

Definition 2.1: *A **mechanism** $\xi = \{(M_1,\ldots,M_n), g(\cdot)\}$ is a collection of n message sets (or strategy sets) and an outcome function $g : M_1 \times \ldots \times M_n \to X$.*

Mechanisms can be considered as procedures for how a collective decision is actually made, including a specification of strategy sets. We can now compare what outcomes are realized in equilibrium through a mechanism with the outcomes we "want" to see happen, as represented by a social choice function. If an outcome function g and a social choice function f coincide on some restriction of their domain, then we say g implements f. The particular restriction depends on the equilibrium concept used.

Definition 2.2: *The mechanism $\xi = \{(M_1,\ldots,M_n), g(\cdot)\}$ **implements** a social choice function f if there is an equilibrium strategy profile $(m_1^*(\cdot),\ldots,m_n^*(\cdot))$ of the game induced by ξ such that $\forall\theta \in \Theta$,*

$$g(m_1^*(\theta_1),\ldots,m_n^*(\theta_n)) = f(\theta_1,\ldots,\theta_n)$$

The term equilibrium here is still vague and we will extend this definition to talk about implementation in various contexts. The mechanism itself might have multiple equilibria. If the outcome of one equilibrium coincides with the scf, then the scf is *partially* implemented. If the outcomes of all equilibria coincide with the scf, then the scf is *fully* implemented. Figure 1 depicts the generic implementation problem.
Due to the revelation principle, we can usually think of the message space as the type space, and most mechanisms we'll look at are direct revelation mechanisms. This also leads us to a notion of incentive compatibility.

Definition 2.3: *A **direct revelation mechanism** for f is a mechanism in which $M_i = \Theta_i$ for all i and $g(\theta) = f(\theta)$ for all $\theta \in \Theta_1 \times \ldots \times \Theta_I$.*

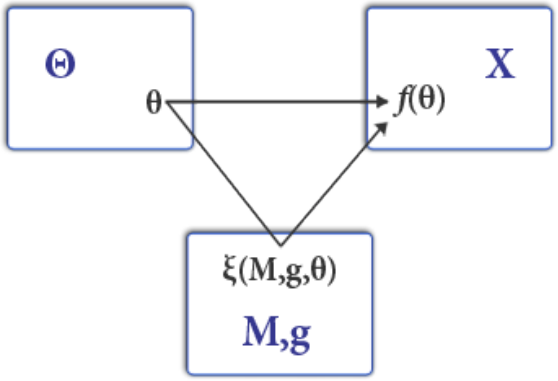

Figure 1: Reiter diagram of a mechanism ξ implementing f

Definition 2.4: *The social choice function f is **truthfully implementable** if the direct revelation mechanism has an equilibrium $(m_1^*(\cdot),\ldots,m_n^*(\cdot))$ in which $m_i^*(\theta_i)=\theta_i$ for all i and θ, i.e. each agent telling the truth is an equilibrium of the induced game. Alternatively, such a mechanism is **incentive compatible**.*

3 Dominant Strategy Implementation

Definition 3.1: *A mechanism $\xi=\{(M_1,\ldots,M_n),g(\cdot)\}$ is a **dominant-strategy implementation** of a social choice function f if there is a dominant strategy equilibrium profile $(m_1^*(\cdot),\ldots,m_I^*(\cdot))$ of the game induced by ξ such that $\forall\theta\in\Theta$, $g(m_1^*(\theta_1),\ldots,m_I^*(\theta_I))=f(\theta_1,\ldots,\theta_I)$.*

Here we start with some scf f and ask whether a mechanism exists such that the outcome from agents playing a dominant strategy matches the scf for all realization of types. If this was the case, then we would have a very robust way of implementing the scf, since we can be confident about what agents will want to do, regardless of what other agents do or what their types are.

Incentive compatibility in this context is also a simple extension.

Definition 3.2: *The scf $f(\cdot)$ is **truthfully implementable in dominant strategies** or **strategy-proof** if for all i and θ_i,*

$$u_i(f(\theta),\theta_i,\theta_{-i})\geq u_i(f(\hat{\theta}_i,\theta_{-i}),\hat{\theta}_i,\theta_{-i})$$

for all $\hat{\theta}_i$ and θ_{-i}.

Now we have the powerful, if nearly trivial, revelation principle for dominant strategy implementation:

Theorem 3.1: *Suppose there exists a mechanism ξ that implements f in dominant strategies. Then f is truthfully implementable in dominant strategies.*

Checking whether an arbitrary scf can be implemented in dominant strategies seems like a monumental task; imagine trying to search over all possible procedures for making a choice. By this theorem though, we know if there is such a mechanism, there is also a direct revelation mechanism where agents report their types honestly. Hence, checking implementability reduces to checking the inequalities in Definition 3.2. This holds for many other equilibrium concepts, which justifies looking only at revelation mechanisms. Of course, this assumes it's practical for agents to fully communicate their types, which could potentially include a full set of beliefs, beliefs about beliefs, etc, without cost, so we might be trading one intractability for another.

4 Nash Implementation

Maskin focuses on general preferences and social choice correspondences, which leads to the following definition:

Definition 4.1: *A mechanism $\{(S_i,\ldots,S_I),g(\cdot)\}$ with $g:\prod S_i\to X$ implements the social choice rule $f:\prod R_i\twoheadrightarrow X$ in Nash equilibrium if*

1. $\forall x \in f(\succeq), \exists (s_1, \ldots, s_I)$ *such that* $x = g(s_1, \ldots, s_I)$ *and* $x \succeq_i g(s_i', s_{-i})$ *for all* i *and all* s_i'.

2. *If* $(s_1, \ldots, s_I)$ *is a Nash equilibrium, then* $g(s) \in f(\succeq)$.

The first part extends our implementation to correspondences, while the second requires that all Nash equilibria result in outcomes in the social choice rule. Then, Maskin gives us the following result:

Theorem 4.1: *If* $f : R^n \twoheadrightarrow X$ *is implementable in Nash equilibrium, then* f *is monotonic.*

This provides an easy first check to see whether a scr is Nash-implementable. However, the converse is not true. We'll need another condition:

Definition 4.2: *A scr* $f : R^n \twoheadrightarrow X$ *satisfies* ***no veto power (NVP)*** *if for all preference profiles* $\succeq$ *and all* $x \in X$*, then for all agents* i *we have*

$$\forall j \neq i, \forall y \in X, x \succeq_j y \implies x \in f(\succeq)$$

Theorem 4.2: *With at least three players, if* f *is monotonic and satisfies NVP, then* f *is Nash-implementable.*

5 Exercises

Exercise 1 (p. 57 of *Toolbox for Economic Design*)**:** *Assume a social choice function $f : \Theta \to X$ is dominant-strategy incentive compatible. Let $\bar{f} : \bar{\Theta} \to X$, where $\bar{\Theta} \subset \Theta$, be the restriction of f to $\bar{\Theta}$, i.e. $\theta \in \bar{\Theta} \implies \bar{f}(\theta) = f(\theta)$. Is $\bar{f}$ also dominant-strategy incentive compatible?*

Exercise 2 (p. 56 of *Toolbox for Economic Design*)**:** *Imagine a society that consists of a single couple, Link and Zelda. The couple is trying to decide whether to paint their castle black, white, or silver. Zelda has one type θ_Z. Her ranking is white $\succ$ silver $\succ$ black. Link, on the other hand, has two possible types: θ_L and $\bar{\theta}_L$. When his type is θ_L, his ranking is black $\succ$ silver $\succ$ white. When he is type $\bar{\theta}_L$, his ranking is silver $\succ$ white $\succ$ black.*

1. *Suppose we wish to implement the following social choice function: $f(\theta_Z, \theta_L) =$ silver and $f(\theta_Z, \bar{\theta}_L) =$ white. Will Link choose to truthfully reveal his type?*
2. *What if the desired social choice function is $f(\theta_Z, \theta_L) =$ black and $f(\theta_Z, \bar{\theta}_L) =$ white?*
3. *What about $f(\theta_Z, \theta_L) =$ silver and $f(\theta_Z, \bar{\theta}_L) =$ silver?*

Exercise 3 (Dutta and Sen (2012), "Nash implementation with partially honest individuals")**:** *Suppose an individual has preferences over which message they send in addition to preferences over outcomes. In particular, suppose individuals don't like to lie when it doesn't help them. Let's call an agent* partially honest *if*

1. *For $a \succ_i b$, the preference over outcome/message pairs are strict: $(a, m_i') \succ_i (b, m_i'')$, $\forall\, m_i', m_i''$*
2. *For $a \sim_i b$ and m_i^* is the honest message, the agent strictly prefers to be honest: $(a, m_i^*) \succ_i (b, m_i')$, $\forall\, m_i' \neq m_i^*$*

Show: If $n \geq 3$ and at least one agents is partially honest, then any SCF satisfying No Veto Power is fully Nash implementable.