Notes on the Revelation Principle[*]
Course Econ 201B

Rodolfo G. Campos
UCLA

February 25, 2005

1 Notation and Definitions

The vector of agents′types $\theta = (\theta_1, .., \theta_n)$ is drawn from the set $\Theta = \Theta_1 \times ... \times \Theta_n$ according to probability density $\phi(\cdot)$. Agent i has expected utility $u_i(x, \theta)$, where $x \in X$, the set of outcomes. A strategy for player i is a function $s_i : \Theta_i \to S_i$, where S_i are the actions available to agent i. A strategy profile is a profile of these functions $s = (s_1, ..., s_n)$ which can also be represented as a single function $s : \Theta \to S_1 \times ... \times S_n$.

Definition 1 *A social choice function is a function $f : \Theta_1 \times ... \times \Theta_n \to X$ that for each possible profile of the agents′ types assigns an outcome $f(\theta_1, .., \theta_n) \in X$.*

A social choice function therefore tells us what outcome is chosen for each configuration of types. We could think that a benevolent planner is the one who does this mapping. This planner is able to look into people′s minds and see their types and choose in the set X in some optimal way (an efficient allocation for example). We don′t ask why $f(\cdot)$ is the social choice function. We just take it as given. What interests us is how we can implement this function when people have private information of their types, and might therefore be tempted to lie.

Definition 2 *A mechanism $\Gamma = (S_1, ..., S_n, g(\cdot))$ is a collection of n strategy spaces and an outcome function $g : S_1 \times ... \times S_n \to X$*

A mechanism is an institution with rules governing the procedure for making the collective choice. S_i denotes the allowed actions for every agent and the function g maps these actions into an outcome.This function differs from $f(\cdot)$ in that We can write the utility functions as functions of g (and therefore, functions of a strategy profile) in the following way:

[*]Everything is based on MWG Chapter 23.

$$\tilde{u}_i\left(s_1, ..., s_n, \theta\right) \equiv u_i\left(g\left(s_1, ..., s_n\right), \theta\right) \tag{1}$$

If we combine a mechanism with the sets of types and their probability distributions we get a Bayesian Game which consists of the elements

$$\left[\left(S_1, ..., S_n\right), \left(\tilde{u}_1\left(s_1, ..., s_n, \theta\right), ..., \tilde{u}_n\left(s_1, ..., s_n, \theta\right)\right), \left(\Theta_1 \times ... \times \Theta_n\right), \phi\left(\cdot\right)\right] \tag{2}$$

Therefore, we can define a Bayesian Nash Equilibrium (BNE) for a mechanism, just as we did for a Bayesian Game.

Definition 3 *The strategy profile $s^* = \left(s_1^*, ..., s_n^*\right)$ is a BNE of mechanism $\Gamma = \left(S_1, ..., S_n, g\left(\cdot\right)\right)$ if for all i and $\theta_i \in \Theta_i$*

$$E_{\theta_{-i}}\left[u_i\left(g\left(s_i^*\left(\theta_i\right), s_{-i}^*\left(\theta_{-i}\right)\right), \theta|\theta_i\right)\right] \geq E_{\theta_{-i}}\left[u_i\left(g\left(\hat{s}_i, s_{-i}^*\left(\theta_{-i}\right)\right), \theta|\theta_i\right)\right] \tag{3}$$

for all $\hat{s}_i \in S_i$.

Definition 4 *The mechanism $\Gamma = \left(S_1, ..., S_n, g\left(\cdot\right)\right)$ implements social choice function $f\left(\cdot\right)$ if there exists a BNE profile s^* of the game induced by Γ such that*

$$g\left(s_1^*\left(\theta_1\right), ..., s_n^*\left(\theta_n\right)\right) = f\left(\theta_1, .., \theta_n\right) \tag{4}$$

for all $\theta \in \Theta$.

Note that this is a weak notion of implementation. The mechanism may have more than one equilibrium. The definition only requires than one of them generates the same outcomes as $f\left(\cdot\right)$.

The issue of finding all social choice functions that are implementable seems difficult, because in principle we would need to consider all possible mechanisms. Fortunately, we can restrict ourselves to direct revelation mechanisms.

Definition 5 *A direct revelation mechanism is a mechanism in which $S_i = \Theta_i$ for all i and $g\left(\theta\right) = f\left(\theta\right)$ for all $\theta \in \Theta$.*

A direct revelation mechanism is a special case of a mechanism where the strategy spaces are simply the type spaces, and thus, each agent's strategy is just announcing a type.

Definition 6 *The social choice function $f\left(\cdot\right)$ is truthfully implementable in Bayesian Nash Equilibrium if $s_i^* = \theta_i$ for all $\theta_i \in \Theta_i$ and all i is a BNE of the direct revelation mechanism $\Gamma = \left(\Theta_1, ..., \Theta_n, f\left(\cdot\right)\right)$. That is, for all i and all $\theta_i \in \Theta_i$*

$$E_{\theta_{-i}}\left[u_i\left(f\left(\theta_i, , \theta_{-i}\right), \theta|\theta_i\right)\right] \geq E_{\theta_{-i}}\left[u_i\left(f\left(\hat{\theta}_i, , \theta_{-i}\right), \theta|\theta_i\right)\right] \tag{5}$$

for all $\hat{\theta}_i \in \Theta_i$.

Note that this again is a weak notion of implementation. There may be more than one BNE. The definition only requires than one of the BNE truthfully implements $f\left(\cdot\right)$.

2 The Revelation Principle

Proposition 7 *Suppose that there exists a mechanism* $\Gamma = (S_1, ..., S_n, g(\cdot))$ *that implements the social choice function* $f(\cdot)$ *in Bayesian Nash Equilibrium. Then* $f(\cdot)$ *is truthfully implementable in Bayesian Nash Equilibrium*

Proof. The proof consists in showing that condition (3) implies condition (5).

Saying that $\Gamma = (S_1, ..., S_n, g(\cdot))$ implements $f(\cdot)$ is saying that there exists a profile of strategies $s^* : \Theta \to S_1 \times ... \times S_n$ such that $g(s^*(\theta)) = f(\theta)$ for all θ, and that for all i and $\theta_i \in \Theta_i$

$$E_{\theta_{-i}}\left[u_i\left(g\left(s_i^*(\theta_i), s_{-i}^*(\theta_{-i})\right), \theta|\theta_i\right)\right] \geq E_{\theta_{-i}}\left[u_i\left(g\left(\hat{s}_i, s_{-i}^*(\theta_{-i})\right), \theta|\theta_i\right)\right] \quad (6)$$

for all $\hat{s}_i \in S_i$.

If the weak inequality holds for all elements of S_i then it must hold if we replace $\hat{s}_i$ by any function $s_i : \Theta_i \to S_i$. The inequality holds for all elements in the domain because for any element in the domain the function is only allowed to assign values in S_i. In particular, we can consider the function s_i^*.

Hence, for all i and $\theta_i \in \Theta_i$

$$E_{\theta_{-i}}\left[u_i\left(g\left(s_i^*(\theta_i), s_{-i}^*(\theta_{-i})\right), \theta|\theta_i\right)\right] \geq E_{\theta_{-i}}\left[u_i\left(g\left(s_i^*\left(\hat{\theta}_i\right), s_{-i}^*(\theta_{-i})\right), \theta|\theta_i\right)\right] \quad (7)$$

for all $\hat{\theta}_i \in \Theta_i$.

Now use the fact that $g(s^*(\theta)) = f(\theta)$ for all configurations of θ and replace $g(\cdot)$ by $f(\cdot)$. Therefore, for all i and $\theta_i \in \Theta_i$

$$E_{\theta_{-i}}\left[u_i\left(f(\theta_i, \theta_{-i}), \theta|\theta_i\right)\right] \geq E_{\theta_{-i}}\left[u_i\left(f\left(\hat{\theta}_i, \theta_{-i}\right), \theta|\theta_i\right)\right] \quad (8)$$

for all $\hat{\theta}_i \in \Theta_i$. This is precisely condition (5) which says that $f(\cdot)$ is truthfully implementable. ■

The Revelation Principle tells us that when when constructing a mechanism (or contract) it is enough to consider mechanisms where the actions are restricted to be announcements of types.

There is one catch. The Revelation principle says that if there is *a* BNE that implements $f(\cdot)$ in the original mechanism then there is *a* BNE that truthfully implements it in a direct revelation mechanism. It does not say that *all* BNE of the direct revelation mechanism implement $f(\cdot)$. Neither does it say that if $f(\cdot)$ is implementable in a direct revelation mechanism then it is implemented in *all* BNE of the original mechanism. Perhaps, the best way of putting it is using the negation.[1]

If $f(\cdot)$ can't be implemented in a direct revelation mechanism, then it can't be implemented in any other mechanism.

[1]Using the fact that $(A \Rightarrow B)$ is equivalent to (not $A \Rightarrow$ not B).