Game Theory

Lecture Notes By

Y. Narahari

Department of Computer Science and Automation

Indian Institute of Science

Bangalore, India

July 2012

Incentive Compatibility and Revelation Theorem

Note: *This is a only a draft version, so there could be flaws. If you find any errors, please do send email to* `hari@csa.iisc.ernet.in`. *A more thorough version would be available soon in this space.*

1 Incentive Compatibility and the Revelation Theorem

The notion of incentive compatibility is perhaps the most fundamental concept in mechanism design, and the revelation theorem is perhaps the most fundamental result in mechanism design. We have already seen that mechanism design involves the preference revelation (or elicitation) problem and the preference aggregation problem. The preference revelation problem involves eliciting truthful information from the agents about their types. In order to elicit truthful information, there is a need to somehow make truth revelation a best response for the agents, consistent with rationality and intelligence assumptions. Offering incentives is a way of doing this; incentive compatibility essentially refers to offering the right amount of incentive to induce truth revelation by the agents. There are broadly two types of incentive compatibility: (1) Truth revelation is a best response for each agent irrespective of what is reported by the other agents; (2) Truth revelation is a best response for each agent whenever the other agents also reveal their true types. The first one is called dominant strategy incentive compatibility (DSIC), and the second one is called Bayesian Nash incentive compatibility (BIC). Since truth revelation is always with respect to types, only direct revelation mechanisms are relevant when formalizing the notion of incentive compatibility. The notion of incentive compatibility was first introduced by Leonid Hurwicz [1].

1.1 Incentive Compatibility (IC)

Definition 1 (Incentive Compatibility) *A social choice function $f : \Theta_1 \times \ldots \times \Theta_n \rightarrow X$ is said to be incentive compatible (or truthfully implementable) if the Bayesian game induced by the direct revelation mechanism $\mathscr{D} = ((\Theta_i)_{i \in N}, f(\cdot))$ has a pure strategy equilibrium $s^*(\cdot) = (s_1^*(\cdot), \ldots, s_n^*(\cdot))$ in which $s_i^*(\theta_i) = \theta_i, \forall \theta_i \in \Theta_i, \forall i \in N$.*

That is, truth revelation by each agent constitutes an equilibrium of the game induced by $\mathscr{D}$. It is easy to infer that if an SCF $f(\cdot)$ is incentive compatible then the direct revelation mechanism

$\mathscr{D} = ((\Theta_i)_{i \in N}, f(\cdot))$ can implement it. That is, directly asking the agents to report their types and using this information in $f(\cdot)$ to get the social outcome will solve both the problems, namely, preference elicitation and preference aggregation.

Based on the type of equilibrium concept used, two types of incentive compatibility are defined.

Definition 2 (Dominant Strategy Incentive Compatibility (DSIC)) *A social choice function $f : \Theta_1 \times \ldots \times \Theta_n \rightarrow X$ is said to be dominant strategy incentive compatible (or truthfully implementable in dominant strategies) if the direct revelation mechanism $\mathscr{D} = ((\Theta_i)_{i \in N}, f(\cdot))$ has a* weakly dominant strategy equilibrium $s^*(\cdot) = (s_1^*(\cdot), \ldots, s_n^*(\cdot))$ *in which $s_i^*(\theta_i) = \theta_i, \forall \theta_i \in \Theta_i, \forall i \in N$.*

That is, truth revelation by each agent constitutes a dominant strategy equilibrium of the game induced by $\mathscr{D}$. Strategy-proof, cheat-proof, straightforward are the alternative phrases used for this property.

Example 1 (Dominant Strategy Incentive Compatibility of Second Price Procurement Auction) It is easy to see that the social choice function implemented by the second price auction is dominant strategy incentive compatible.

Using the definition of a dominant strategy equilibrium in Bayesian games (Section **??**), the following necessary and sufficient condition for an SCF $f(\cdot)$ to be dominant strategy incentive compatible can be easily derived:

$$u_i\left(f\left(\theta_i, \theta_{-i}\right), \theta_i\right) \geq u_i(f(\hat{\theta_i}, \theta_{-i}), \theta_i), \forall i \in N, \forall \theta_i \in \Theta_i, \forall \theta_{-i} \in \Theta_{-i}, \forall \hat{\theta_i} \in \Theta_i. \tag{1}$$

The above condition says that if the SCF $f(.)$ is DSIC, then, irrespective of what the other agents report, it is always a best response for agent i to report his true type θ_i.

Definition 3 (Bayesian Incentive Compatibility (BIC)) *A social choice function $f : \Theta_1 \times \ldots \times \Theta_n \rightarrow X$ is said to be Bayesian incentive compatible (or truthfully implementable in Bayesian Nash equilibrium) if the direct revelation mechanism $\mathscr{D} = ((\Theta_i)_{i \in N}, f(\cdot))$ has a* Bayesian Nash equilibrium $s^*(\cdot) = (s_1^*(\cdot), \ldots, s_n^*(\cdot))$ *in which $s_i^*(\theta_i) = \theta_i, \forall \theta_i \in \Theta_i, \forall i \in N$.*

That is, truth revelation by each agent constitutes a Bayesian Nash equilibrium of the game induced by $\mathscr{D}$.

Example 2 (Bayesian Incentive Compatibility of First Price Procurement Auction) We have seen that the first price procurement auction for a single indivisible item implements the following social choice function:

$$f(\theta) = (y_0(\theta), y_1(\theta), y_2(\theta), t_0(\theta), t_1(\theta), t_2(\theta))$$

with

$$\begin{array}{rcl} y_0(\theta) & = & 0 \quad \forall \theta \in \Theta \\ y_1(\theta) & = & 1 \quad \text{if } \theta_1 \leq \theta_2 \\ & = & 0 \quad \text{otherwise} \\ y_2(\theta) & = & 1 \quad \text{if } \theta_1 > \theta_2 \\ & = & 0 \quad \text{otherwise} \\ t_1(\theta) & = & \dfrac{1+\theta_1}{2} y_1(\theta) \\ t_2(\theta) & = & \dfrac{1+\theta_2}{2} y_2(\theta) \end{array}$$

$$t_0(\theta) \quad = \quad -(t_1(\theta) + t_2(\theta)).$$

If seller 1 has type θ_1, then his optimal bid $\hat{\theta}_1$ is obtained by solving

$$\max_{\hat{\theta}_1} \left(\frac{1 + \hat{\theta}_1}{2} - \theta_1 \right) P\{\theta_2 \geq \hat{\theta}_1\}.$$

This is the same as

$$\max_{\hat{\theta}_1} \left(\frac{1 + \hat{\theta}_1}{2} - \theta_1 \right) (1 - \hat{\theta}_1).$$

This yields $\hat{\theta}_1 = \theta_1$. Thus it is optimal for seller 1 to reveal his true private value if seller 2 reveals his true value. The same situation applies to seller 2. This implies that the social choice function is Bayesian Nash incentive compatible (since the equilibrium involved is a Bayesian Nash equilibrium).

Using the definition of a Bayesian Nash equilibrium in Bayesian games (Section **??**), the following necessary and sufficient condition for an SCF $f(\cdot)$ to be Bayesian incentive compatible can be easily derived:

$$E_{\theta_{-i}} \left[u_i \left(f \left(\theta_i, \theta_{-i} \right), \theta_i \right) | \theta_i \right] \geq E_{\theta_{-i}} [u_i(f(\hat{\theta}_i, \theta_{-i}), \theta_i) | \theta_i], \forall i \in N, \forall \theta_i \in \Theta_i, \forall \hat{\theta}_i \in \Theta_i \tag{2}$$

where the expectation is taken over the type profiles of agents other than agent i.

Note 1 *If a social choice function $f(\cdot)$ is dominant strategy incentive compatible then it is also Bayesian incentive compatible. The proof of this follows trivially from the fact that a weakly dominant strategy equilibrium is necessarily a Bayesian Nash equilibrium.*

1.2 The Revelation Principle for Dominant Strategy Equilibrium

The revelation principle basically illustrates the relationship between an indirect mechanism $\mathscr{M}$ and a direct revelation mechanism $\mathscr{D}$ with respect to a given SCF $f(\cdot)$. This result enables us to restrict our inquiry about truthful implementation of an SCF to the class of direct revelation mechanisms only.

Theorem 1 *Suppose that there exists a mechanism $\mathscr{M} = (S_1, \ldots, S_n, g(\cdot))$ that implements the social choice function $f(\cdot)$ in dominant strategy equilibrium. Then $f(\cdot)$ is dominant strategy incentive compatible.*

Proof: If $\mathscr{M} = (S_1, \ldots, S_n, g(\cdot))$ implements $f(\cdot)$ in dominant strategies, then there exists a profile of strategies $s^*(\cdot) = (s_1^*(\cdot), \ldots, s_n^*(\cdot))$ such that

$$g\left(s_1^*(\theta_1), \ldots, s_n^*(\theta_n)\right) = f\left(\theta_1, \ldots, \theta_n\right) \ \forall \left(\theta_1, \ldots, \theta_n\right) \in \Theta \tag{3}$$

and

$$\begin{aligned} u_i(g(s_i^*(\theta_i), s_{-i}(\theta_{-i})), \theta_i) \quad &\geq \quad u_i(g(s_i^{'}(\theta_i), s_{-i}(\theta_{-i})), \theta_i) \\ &\forall i \in N, \forall \theta_i \in \Theta_i, \forall \theta_{-i} \in \Theta_{-i}, \forall s_i^{'}(\cdot) \in S_i, \forall s_{-i}(\cdot) \in S_{-i}. \end{aligned} \tag{4}$$

Condition (4) implies, in particular, that

$$\begin{aligned} u_i(g(s_i^*(\theta_i), s_{-i}^*(\theta_{-i})), \theta_i) \quad &\geq \quad u_i(g(s_i^*(\hat{\theta}_i), s_{-i}^*(\theta_{-i})), \theta_i) \\ &\forall i \in N, \forall \theta_i \in \Theta_i, \forall \hat{\theta}_i \in \Theta_i, \forall \theta_{-i} \in \Theta_{-i}. \end{aligned} \tag{5}$$

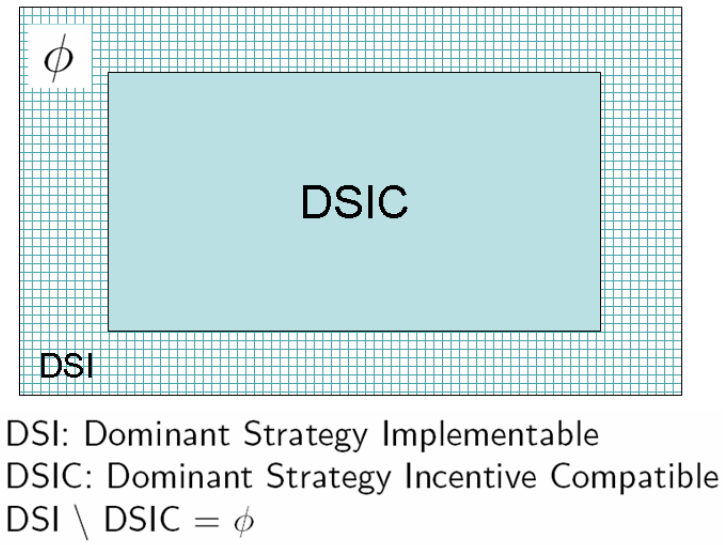


Figure 1: Revelation principle for dominant strategy equilibrium

Conditions (3) and (5) together imply that

$$u_i\left(f\left(\theta_i, \theta_{-i}\right), \theta_i\right) \geq u_i(f(\hat{\theta}_i, \theta_{-i}), \theta_i), \forall i \in N, \forall \theta_i \in \Theta_i, \forall \theta_{-i} \in \Theta_{-i}, \forall \hat{\theta}_i \in \Theta_i.$$

But this is precisely condition (1), the condition for $f(\cdot)$ to be truthfully implementable in dominant strategies.

Q.E.D.

The idea behind the revelation principle can be understood with the help of Figure 1. In this picture, **DSI** represents the set of all social choice functions that are implementable in dominant strategies and **DSIC** is the set of all social choice functions that are dominant strategy incentive compatible. The picture depicts the obvious fact that **DSIC** is a subset of **DSI** and illustrates the revelation theorem by showing that the set difference between these two sets is the empty set, thus implying that **DSIC** is precisely the same as **DSI**.

1.3 The Revelation Principle for Bayesian Nash Equilibrium

Theorem 2 *Suppose that there exists a mechanism $\mathscr{M} = (S_1, \ldots, S_n, g(\cdot))$ that implements the social choice function $f(\cdot)$ in Bayesian Nash equilibrium. Then $f(\cdot)$ is truthfully implementable in Bayesian Nash equilibrium (Bayesian incentive compatible).*

Proof: If $\mathscr{M} = (S_1, \ldots, S_n, g(\cdot))$ implements $f(\cdot)$ in Bayesian Nash equilibrium, then there exists a profile of strategies $s^*(\cdot) = (s_1^*(\cdot), \ldots, s_n^*(\cdot))$ such that

$$g\left(s_1^*(\theta_1), \ldots, s_n^*(\theta_n)\right) = f\left(\theta_1, \ldots, \theta_n\right) \ \forall \left(\theta_1, \ldots, \theta_n\right) \in \Theta \tag{6}$$

and

$$\begin{aligned} E_{\theta_{-i}}\left[u_i(g(s_i^*(\theta_i), s_{-i}^*(\theta_{-i})), \theta_i)|\theta_i\right] \quad &\geq \quad E_{\theta_{-i}}\left[u_i(g(s_i^{'}(\theta_i), s_{-i}^*(\theta_{-i})), \theta_i)|\theta_i\right] \\ &\forall i \in N, \forall \theta_i \in \Theta_i, \forall s_i^{'}(\cdot) \in S_i. \end{aligned} \tag{7}$$

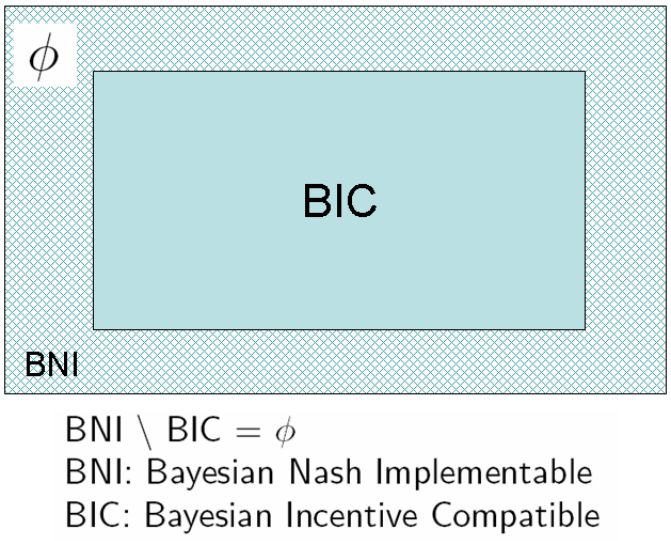


Figure 2: Revelation principle for Bayesian Nash equilibrium

Condition (7) implies, in particular, that

$$E_{\theta_{-i}}\left[u_i(g(s_i^*(\theta_i), s_{-i}^*(\theta_{-i})), \theta_i)|\theta_i\right] \geq E_{\theta_{-i}}\left[u_i(g(s_i^*(\hat{\theta}_i), s_{-i}^*(\theta_{-i})), \theta_i)|\theta_i\right] \quad \forall i \in N, \forall \theta_i \in \Theta_i, \forall \hat{\theta}_i \in \Theta_i. \tag{8}$$

Conditions (6) and (8) together imply that

$$E_{\theta_{-i}}\left[u_i\left(f\left(\theta_i, \theta_{-i}\right), \theta_i\right)|\theta_i\right] \geq E_{\theta_{-i}}\left[u_i(f(\hat{\theta}_i, \theta_{-i}), \theta_i)|\theta_i\right], \forall i \in N, \forall \theta_i \in \Theta_i, \forall \hat{\theta}_i \in \Theta_i.$$

But this is precisely condition (2), the condition for $f(\cdot)$ to be truthfully implementable in Bayesian Nash equilibrium.

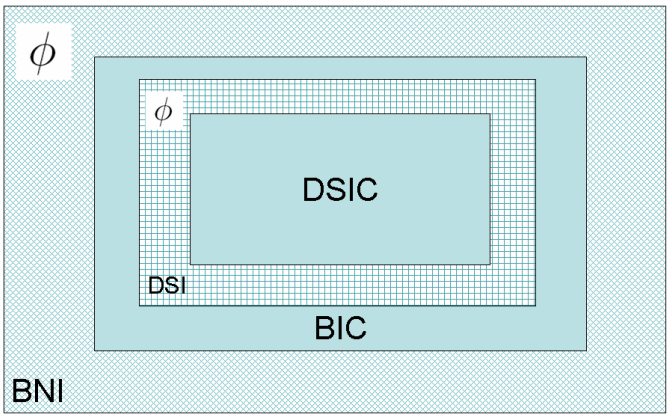
Q.E.D.

In a way similar to the revelation principle for dominant strategy equilibrium, the revelation principle for Bayesian Nash equilibrium can be explained with the help of Figure 2. In this picture, **BNI** represents the set of all social choice functions which are implementable in Bayesian Nash equilibrium and **BIC** is the set of all social choice functions which are Bayesian incentive compatible. The picture depicts the fact that **BIC** is a subset of **BNI** and illustrates the revelation theorem by showing that the set difference between these two sets is the empty set, thus implying that **BIC** is precisely the same as **BNI**.

Figure 3 provides a combined view of both the revelation theorems that we have seen in this section.

References

[1] L. Hurwicz. On informationally decentralized systems. In C.B. McGuire and R. Radner, editors, *Decision and Organization*. North-Holland, Amsterdam, 1972.



BNI \ BIC = ϕ
BNI: Bayesian Nash Implementable
BIC: Bayesian Incentive Compatible

DSI: Dominant Strategy Implementable
DSIC: Dominant Strategy Incentive Compatible
DSI \ DSIC = ϕ

Figure 3: Combined view of revelation theorems for dominant strategy equilibrium and Bayesian Nash equilibrium