DDA6205 Spring 2024 - Assignment 3

Yohandi

May 14, 2024

The Gibbard-Satterthwaite Theorem

A decision rule d is dominant strategy incentive compatible (or strategy-proof) if the social choice function $f = (d, t^0)$ is dominant strategy incentive compatible, where t^0 is the transfer function that is identically 0. A decision rule d is dictatorial if there exists i such that $d(\theta) \in \arg\max_{d \in R_d} v_i(d, \theta_i)$ for all θ, where $R_d = \{d \in D \mid \exists \theta \in \Theta : d = d(\theta)\}$ is the range of d.

Theorem 1 Suppose that D is finite and type spaces include all possible strict orderings over D. A decision rule with at least three elements in its range is dominant strategy incentive compatible (strategy-proof) if and only if it is dictatorial.

Prove the theorem.

Definitions:

- Let Θ denote the set of all profiles of strict rankings over D.
- $d : \Theta \to D$ is the decision rule.
- A decision rule d is strategy-proof if no agent can benefit by unilaterally misreporting their preferences, assuming all other agents report honestly.
- A decision rule d is dictatorial if there exists an agent i such that for all $\theta \in \Theta$, $d(\theta) = \theta_i(1)$, where $\theta_i(1)$ is the top choice of agent i.

Given D with $|D| \geq 3$, we are to prove that a decision rule $d : \Theta \to D$ is strategy-proof if and only if it is dictatorial.

($\Rightarrow$) Strategy-proof implies Dictatorial:

For the sake of contradiction, assume that d is strategy-proof but not dictatorial. Define Θ as the set of all preference profiles over a finite set of alternatives D with $|D| \geq 3$.

Consider any profile $\theta \in \Theta$ where a particular alternative B is ranked last by every voter due to the unanimity principle. Despite B being least preferred universally, assume that $d(\theta) \neq B$, which is consistent with the strategy-proof assumption of d.

- Modify θ by promoting B sequentially in each voter's ranking. This is done by shifting B one position higher in the ranking of one voter at a time, keeping other voters' rankings constant during each step.
- Maintain the strategy-proof nature of d, which dictates that the decision outcome $d(\theta')$ must remain consistent (i.e., $K \neq B$) as long as B is not the top choice of any voter.

Identify a pivotal voter r through the following process by continue promoting B until $d(\theta') = B$, where θ' reflects B moved to the top position in a single voter's (say voter r's) ranking while others still rank B lower.

Voter	Initial Ranking	Modified Ranking (If B raised by r)
1	$B > \ldots > A$	$B > \ldots > A$
2	$B > \ldots > A$	$B > \ldots > A$
$\vdots$	$\vdots$	$\vdots$
r	$K > A > \ldots > B$	$B > K > A > \ldots$
$\vdots$	$\vdots$	$\vdots$
n	$C > \ldots > B$	$C > \ldots > B$

Observe that r is a pivotal voter for B, implying that a change in r's report from B not being at the top to B at the top results in B being selected by d. The preceding contradicts the assumption of strategy-proofness, as r can manipulate the outcome by changing their ranking of B, directly influencing $d(\theta')$ to select B.

To further substantiate the role of r, consider rotating the roles of B with other alternatives C and K, ensuring each undergoes a similar profile manipulation:

- For each alternative $X \in D \backslash \{B\}$, repeat the ranking modification process. Elevate X in the rankings under the same conditions and observe if a similar pivotal shift occurs, pointing r as the determining voter for X.

- If for all such X, voter r is observed to dictate the outcome when they rank X at the top, it provides a clear pattern of dictatorial control by r across multiple alternatives, reinforcing the need for d to be dictated by r's preferences.

The impact of individual voter preference manipulations on the outcome dictated by rule d. The following matrix illustrates the shifts in voter preferences, particularly focusing on voter r's ability to change the collective decision simply by changing their top preference.

Voter	Initial Profile	Modified Profile	Final Profile
1	$K > \ldots$	$B > K > \ldots$	$K > B > \ldots$
2	$\ldots > K$	$\ldots > B > K$	$\ldots > K > B$
$\vdots$	$\vdots$	$\vdots$	$\vdots$
r	$B > \ldots > K$	$K > B > \ldots$	$B > K > \ldots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$
n	$\ldots > B$	$\ldots > K$	$\ldots > B$

Therefore, the preceding demonstrates that d is susceptible to being controlled by a single voter's top choice, thereby implying a dictatorial decision rule. By contradiction, if d is strategy-proof for $|D| \geq 3$, then it must necessarily be dictatorial. By having $|D| \geq 3$ and rankings completion ensures that without dictatorship, manipulations by voters would lead to a different decision by d, violating strategy-proofness.

($\Leftarrow$) Dictatorial implies Strategy-proof:

Assume d is dictatorial, controlled by dictator i. This means that for any profile $\theta \in \Theta$, where Θ is the set of all preference profiles over a finite set of alternatives D with $|D| \geq 3$, the outcome $d(\theta)$ is always determined by $\theta_i(1)$, the top choice of i at profile θ.

Consider the following arguments:

- For any agent $j \neq i$, examine the impact of any unilateral deviation in their reported preference. Suppose j changes their reported preference from θ_j to θ_j' while all other agents' preferences remain fixed.

- Since d is dictated by θ_i alone, the decision outcome remains $d(\theta) = \theta_i(1)$ regardless of the change in j's report. Therefore, the decision outcome as determined by $d(\theta')$ is the same as $d(\theta)$, where $\theta' = (\theta_j', \theta_{-j})$.

, or, mathematically, the preceding can be expressed as $d(\theta_j', \theta_{-j}) = d(\theta_j, \theta_{-j}) = \theta_i(1)$, which implies that j's deviation has no impact on the outcome.

Now consider the dictator i:

- The dictator i always achieves their most preferred outcome under truthful reporting, as $d(\theta) = \theta_i(1)$. Suppose i considers a deviation to a different report θ_i'.

- Under the dictatorship, any report by i results in the outcome being i's top choice from the new report, so $d(\theta_i', \theta_{-i}) = \theta_i'(1)$. However, since $\theta_i(1)$ is i's genuine top choice under the original profile, there is no incentive for i to misreport because $\theta_i(1)$ maximizes i's preference.

Therefore, no agent can influence the outcome of d through unilateral deviation from their true preferences. Non-dictators cannot change the outcome at all, and the dictator achieves the optimal result without misrepresentation. Consequently, d is inherently strategy-proof because unilateral deviations from honest reporting do not benefit any voter, preserving the integrity of the decision rule under strategy-proofness.

By both sufficiency and necessity, a decision rule with at least three elements in its range is dominant strategy incentive compatible (strategy-proof) if and only if it is dictatorial.

Groves' Schemes

Theorem 2 *(I)* If d be an efficient decision rule and for each i there exists a function $x_i : X_{-i} \to \mathbb{R}$ such that

$$t_i(\theta) = x_i(\theta_{-i}) + \sum_{j \neq i} v_j(d(\theta), \theta_j),$$

then (d, t) is dominant strategy incentive compatible.

(II) Conversely, if d is an efficient decision rule, (d, t) is dominant strategy incentive compatible, and the type spaces are complete in the sense that $\{(v_i(\cdot, \theta_i) | \theta_i \in \Theta_i\} = \{v : D \to \mathbb{R}\}$ for each i, then for each i there exists a function $x_i : \times_{j \neq i} \Theta_j \to \mathbb{R}$ such that the transfer function t_i satisfies

$$t_i(\theta) = x_i(\theta_{-i}) + \sum_{j \neq i} v_j(d(\theta), \theta_j).$$

Prove the theorem.

Proof of Theorem 2 *(I)*:

Let us first show *(I)*. Suppose to the contrary that d is an efficient decision rule and for each i there exists a function $x_i : X_{-i} \to \mathbb{R}$ such that the transfer function t_i satisfies

$$t_i(\theta) = x_i(\theta_{-i}) + \sum_{j \neq i} v_j(d(\theta), \theta_j),$$

but that (d, t) is not dominant strategy incentive compatible. Then there exists i, θ, θ'_i such that

$$v_i(d(\theta_{-i}, \theta'_i), \theta_i) + t_i(\theta_{-i}, \theta'_i) > v_i(d(\theta), \theta_i) + t_i(\theta).$$

From this it implies that

$$v_i(d(\theta_{-i}, \theta'_i), \theta_i) + x_i(\theta_{-i}) + \sum_{j \neq i} v_j(d(\theta_{-i}, \theta'_i), \theta_j) > v_i(d(\theta), \theta_i) + x_i(\theta_{-i}) + \sum_{j \neq i} v_j(d(\theta), \theta_j),$$

or that

$$v_i(d(\theta_{-i}, \theta'_i), \theta_i) + \sum_{j \neq i} v_j(d(\theta_{-i}, \theta'_i), \theta_j) > v_i(d(\theta), \theta_i) + \sum_{j \neq i} v_j(d(\theta), \theta_j).$$

This contradicts the efficiency of d and so our supposition was incorrect.

Proof of Theorem 2 *(II)*:

Suppose d is an efficient decision rule and (d, t) is dominant strategy incentive compatible, with complete type spaces, i.e., each $v_i(\cdot, \theta_i)$ is a function from $D \to \mathbb{R}$. For each agent i, if x_i is independent of θ_i, then there exists a function $x_i : \Theta \to \mathbb{R}$ such that the transfer function t_i satisfies:

$$t_i(\theta) = x_i(\theta) + \sum_{j \neq i} v_j(d(\theta), \theta_j)$$

To show the preceding, let's assume for contradiction that there exists θ and θ'_{-i} such that $x_i(\theta) \neq x_i(\theta_{-i}, \theta'_i)$, that implies x_i depends on θ_i. By dominant strategy incentive compatibility, for any $\theta_i, \theta'_i \in \Theta_i$, the utility for agent i when they report truthfully should be at least as good as when they misreport. Mathematically, for all $\theta_{-i} \in \times_{j \neq i} \Theta_j$:

$$
\begin{aligned}
v_i(d(\theta), \theta_i) + t_i(\theta) &\geq v_i(d(\theta_{-i}, \theta'_i), \theta_i) + t_i(\theta_{-i}, \theta'_i) \\
v_i(d(\theta), \theta_i) + x_i(\theta_{-i}) + \sum_{j \neq i} v_j(d(\theta), \theta_j) &\geq v_i(d(\theta_{-i}, \theta'_i), \theta_i) + x_i(\theta_{-i}) + \sum_{j \neq i} v_j(d(\theta_{-i}, \theta'_i), \theta_j) \qquad (t_i\text{'s expansion})
\end{aligned}
$$

Let's also introduce a small perturbation $\epsilon(d(\theta_{-i}, \theta'_i)) = \frac{1}{2} |x_i(\theta) - x_i(\theta_{-i}, \theta'_i)|$. Then, we require:

$$
\begin{aligned}
v_i(d(\theta_{-i}, \theta'_i), \omega_i) + \sum_{j \neq i} v_j(d(\theta_{-i}, \theta'_i), \theta_j) &= \epsilon(d(\theta_{-i}, \theta'_i)) \\
\Rightarrow v_i(d(\theta), \omega_i) + \sum_{j \neq i} v_j(d(\theta), \theta_j) &= \epsilon(d(\theta)) = 0, \forall d(\theta) \neq d(\theta_{-i}, \theta'_i)
\end{aligned}
$$

for some ω_i in the set of Θ_i. Along with the efficiency of d, $d(\theta_{-i}, \omega_i) = d(\theta_{-i}, \theta'_i)$ must also be true. Consequently, the utility to i from the corresponding truthful and dishonest announcements at ω_i must be:

- $u_i(\theta_{-i}, \theta_i, \underbrace{\theta_i}_{\text{truthful}}, d, t) = \epsilon(d(\theta_{-i}, \theta'_i)) + x_i(\theta_{-i}, \theta'_i)$ for truthful announcement.
- $u_i(\theta_{-i}, \theta_i, \underbrace{\omega_i}_{\text{dishonest}}, d, t) = x_i(\theta)$ for dishonest announcement.

This contradicts the assumption of dominant strategy incentive compatibility since $u_i(\theta_{-i}, \theta_i, \omega_i, d, t) > u_i(\theta_{-i}, \theta_i, \theta_i, d, t)$; thus $x_i(\theta_i)$ cannot depend on θ_i. Since our assumption was incorrect, x_i must be independent of θ_i; therefore, it is defined only over θ_{-i}, confirming the theorem's statement:

$$
t_i(\theta) = x_i(\theta_{-i}) + \sum_{j \neq i} v_j(d(\theta), \theta_j), \forall i
$$