randomly selecting a dictator, serial dictatorship

0.0.20 Digression: The Revelation Principle

The Revelation Principle originated in Gibbard’s proof of the Gibbard-Satterthwaite Theorem.

Notation for a game: agent $i$’s strategy set is $S_i$, $\sigma_i : \mathcal{L} \rightarrow S_i$ denotes a strategy of player $i$ (i.e., a choice based upon his preferences over the alternatives $A$), and

$$\tau : \prod_{i=1}^{N} S_i \rightarrow A$$

denotes the outcome function of the game (i.e., how the game determines an alternative based upon the strategic choices of the players). A game or mechanism is denoted

$$\mathcal{L}^N, \tau \circ \sigma.$$

Letting $\sigma : \mathcal{L}^N \rightarrow \prod_{i=1}^{N} S_i$ be defined by

$$\sigma = (\sigma_1, \ldots, \sigma_N),$$

the social choice function $f$ is implemented by the strategy profile $(\sigma_i)_{1 \leq i \leq N}$ in the game $(\prod S_i, \tau)$ if

$$\tau \circ \sigma = f,$$

i.e., $f$ results when the agents employ the strategies $(\sigma_i)_{1 \leq i \leq N}$.

A revelation game or mechanism is a game in which each $S_i = \mathcal{L}$ (i.e., each agent is given the opportunity to report his complete ranking of the alternatives in $A$).

**Theorem 14 (Revelation Principle for Dominant Strategies)** If $(\sigma_i)_{1 \leq i \leq N}$ is a dominant strategy equilibrium in the game $(\prod S_i, \tau)$, then honest revelation by each agent is a dominant strategy equilibrium in the game $(\mathcal{L}^N, \tau \circ \sigma)$.

**Proof.** We need to show that for each agent $i$ and every $(L_1, \ldots, L_i, \ldots, L_N)$, and every $L_i$,

$$\tau \circ \sigma(L_1, \ldots, L_i, \ldots, L_N) \geq_{L_i} \tau \circ \sigma(L_1, \ldots, L'_i, \ldots, L_N)$$

We have

$$\tau \circ \sigma(L_1, \ldots, L_i, \ldots, L_N) = \tau(\sigma_1(L_1), \ldots, \sigma_i(L_i), \ldots, \sigma_N(L_N)) \geq L_i, \tau(\sigma_1(L_1), \ldots, \sigma_i(L'_i), \ldots, \sigma_N(L_N)) = \tau \circ \sigma(L_1, \ldots, L'_i, \ldots, L_N),$$

where the inequality is true because $(\sigma_i)_{1 \leq i \leq N}$ is a dominant strategy equilibrium. 

A verbal explanation of the revelation principle is as follows. Suppose we are given
the dominant strategy equilibrium \((\sigma_i)_{1 \leq i \leq N}\) in the game \((\prod S_i, \tau)\). Imagine that we ask each agent \(i\), "Report to me (as a nonstrategic and honest "operator" of the game) your preferences and I will carry out for you in the game \((\prod S_i, \tau)\) the action you would take according to the strategy \(\sigma_i\)." The claim is that it is a dominant strategy for agent \(i\) to be honest in his report, for if he had some reason to lie for some profile of reported preferences of the other players, then he would also have reason to not use \(\sigma_i\) in the game \((\prod S_i, \tau)\).

0.0.20.1 The Meaning of the Revelation Principle

The implication of the revelation principle is that we can examine all social choice functions that result from dominant strategy equilibria of all possible games simply by considering those that result from honest revelation being a dominant strategy in a revelation mechanism. In other words, the set of all strategy-proof social choice functions consists of all social choice functions that can be implemented as dominant strategy equilibria of arbitrary games. This is important because we may not want to be obsessed with truth-telling as an end in itself, or solely with revelation games.

**Theorem 15 (Gibbard-Satterthwaite Theorem)** If \(A\) has at least 3 elements and \(f : \mathcal{L}^N \to A\) is onto and strategy-proof, then \(f\) is dictatorial.

**Corollary 16** If \(A\) has at least 3 elements and \(f : \mathcal{L}^N \to A\) is Pareto optimal and is implemented by a dominant strategy equilibrium of some game \((\prod S_i, \tau)\), then \(f\) is dictatorial.

**Example 49** Consider the following two person game:

<table>
<thead>
<tr>
<th>T/2</th>
<th>L</th>
<th>R</th>
</tr>
</thead>
<tbody>
<tr>
<td>T</td>
<td>(\frac{\theta_1}{2}, \frac{\theta_2}{2})</td>
<td>(1 - \theta_1, 1 - \theta_2)</td>
</tr>
<tr>
<td>B</td>
<td>(\frac{1 - \theta_1}{2}, \frac{1 - \theta_2}{2})</td>
<td>(\frac{\theta_1}{2}, \frac{\theta_2}{2})</td>
</tr>
</tbody>
</table>

There are four possible outcomes: \((T, L), (T, R), (B, L), (B, R)\). Here, \(\theta_1 \in [0, 1]\) (player 1’s type) is a parameter that player 1 privately observes that determines his preferences over the four outcomes, while \(\theta_2\) (player 2’s type) is a parameter that player 2 privately observes that determines his preferences over the four outcomes.

A strategy for each player is an function that specifies an action for each of his possible types. Notice that the following strategies are dominant for each of the two players:

player 1 :
\[
\begin{cases} 
T & \text{if } \theta_1 \geq \frac{1}{2} \\
B & \text{if } \theta_1 < \frac{1}{2}
\end{cases}
\]

player 2 :
\[
\begin{cases} 
R & \text{if } \theta_2 \geq \frac{1}{2} \\
L & \text{if } \theta_2 < \frac{1}{2}
\end{cases}
\]
Now consider the revelation game specified by the outcome function $f$:

$$
f(\theta_1, \theta_2) = \begin{cases} 
(T, L) & \text{if } \theta_1 \geq \frac{1}{2} \text{ and } \theta_2 < \frac{1}{2}; \\
(T, R) & \text{if } \theta_1 \geq \frac{1}{2} \text{ and } \theta_2 \geq \frac{1}{2}; \\
(B, L) & \text{if } \theta_1 < \frac{1}{2} \text{ and } \theta_2 < \frac{1}{2}; \\
(B, R) & \text{if } \theta_1 < \frac{1}{2} \text{ and } \theta_2 \geq \frac{1}{2}.
\end{cases}
$$

This is derived by considering the outcome as a function of $\theta_1$ and $\theta_2$ in the above game when the two players play their dominant strategies. Can we verify that this game is strategy-proof, i.e., it is a dominant strategy for each player to report honestly? Consider player 1 when $\theta_1$ is his type and he considers reporting $\theta_1^*$:

- if $\theta_1 < \frac{1}{2}$, then reporting any $\theta_1^* < \frac{1}{2}$ doesn’t benefit player 1;
- if $\theta_1 < \frac{1}{2}$, then player 1 receives $1/2$ when player 2 plays $L$ and $1 - \theta_1$ when 2 plays $R$. These payoffs reduce to $\theta_1 < 1/2$ and $1/2 < 1 - \theta_1$, respectively, if he lies and instead reports $\theta_1^* \geq \frac{1}{2};$
- if $\theta_1 \geq \frac{1}{2}$, then reporting any $\theta_1^* \geq \frac{1}{2}$ doesn’t benefit player 1;
- if $\theta_1 \geq \frac{1}{2}$, then player 1 receives $\theta_1$ when player 2 plays $L$ and $1/2$ when 2 plays $R$. These payoffs reduce to $1/2 \leq \theta_1$ and $1 - \theta_1 \leq 1/2$, respectively, if he lies and instead reports $\theta_1^* \geq \frac{1}{2}.$

We thus see that player 1 cannot profit by misreporting his type. A similar analysis applies to the decision problem of player 2.

We have replaced a simple, $2 \times 2$ game with a game in which each player reports his type. The incentives of the original game and its equilibrium outcome as a function of the pair of types has been preserved. Is the revelation game "better" or "simpler" in some sense? I don't think so, but that's not the point. The point is that we can study all possible dominant strategy outcomes of all possible games by studying the more limited class of revelation games in which honest reporting is the dominant strategy of each player.

### 0.0.21 Monotonicity

This is a property that we need to discuss as a prelude to proving the Gibbard-Satterthwaite Theorem. It also has applications beyond this theorem.

**Monotonic:** $f$ is monotonic if it satisfies the following property for all $f(L_1, \ldots, L_i, \ldots, L_N) = a$: Suppose that in moving from the profile $(L_1, \ldots, L_i, \ldots, L_N)$ to $(L'_1, \ldots, L'_i, \ldots, L'_N)$ the alternative $a$ does not fall in the ranking relative to any other alternative $b$. Then $f(L'_1, \ldots, L'_i, \ldots, L'_N) = a$.

**Monotonicity of a correspondence $f: L \rightarrow 2^A$:** $f$ is monotonic if it satisfies the following property for all $a \in f(L_1, \ldots, L_i, \ldots, L_N)$: Suppose that in moving from the profile $(L_1, \ldots, L_i, \ldots, L_N)$ to $(L'_1, \ldots, L'_i, \ldots, L'_N)$ the alternative $a$ does not fall in the ranking relative to any other alternative $b$. Then $a \in f(L'_1, \ldots, L'_i, \ldots, L'_N)$.

Monotonicity is the key property in the sense that it concerns the value of $f$ at distinct profiles. If $a$ is the "correct" choice when preferences are given by $(L_1, \ldots, L_i, \ldots, L_N)$, and $a$ can only move up in the ranking as preferences change to $(L_1, \ldots, L_i, \ldots, L_N)$,
then shouldn’t α be the choice when preferences are given by \((L'_{1}, \ldots, L'_{i}, \ldots, L'_{N})\)? The purpose of this work is to explore reasonable properties that a social choice function \(\phi\) should have. Is monotonicity reasonable?

Monotonicity will be shown to be closely related to various notions of incentive compatibility. Roughly, this means that social choice functions that are implemented when agents act strategically are necessarily monotonic. We thus may not be able to avoid monotonicity if we believe that agents are allowed to pursue their individual self-interests. Notice that \(x_{i}\) continues to be optimal for agent \(i\) subject to his budget constraint after the preferences are changed from \((L_{1}, \ldots, L_{i}, \ldots, L_{N})\) to \((L'_{1}, \ldots, L'_{i}, \ldots, L'_{N})\). The allocation \(\alpha = (x_{i})_{1 \leq i \leq N}\) is thus a Walrasian equilibrium allocation for both \((L_{1}, \ldots, L_{i}, \ldots, L_{N})\) and \((L'_{1}, \ldots, L'_{i}, \ldots, L'_{N})\), which verifies that this correspondence is monotonic.

**Example 50** Let’s consider the Borda Count with 3 agents and 3 alternatives. \(F^{BC}\) is a correspondence. Let’s demonstrate with an example that it is not monotonic. Consider the following preferences:

\[
\begin{align*}
L_{1} & : a > b > c \\
L_{2} & : b > c > a \\
L_{3} & : c > a > b
\end{align*}
\]

We have \(F^{BC}(L_{1}, L_{2}, L_{3}) = \{a, b, c\}\) (each alternative receives 6 votes). Now consider \(L_{2} : b > a > c\). We have \(F^{BC}(L_{1}, L'_{2}, L_{3}) = \{a\}\). The alternative \(b\) does not fall in any agent’s ranking in moving from \((L_{1}, L_{2}, L_{3})\) to \((L_{1}, L'_{2}, L_{3})\), and yet \(b \notin F^{BC}(L_{1}, L'_{2}, L_{3})\). Notice again that when we need an example concerning incentives to vote strategically, we turn to the Condorcet triple.

**Example 51** We consider the correspondence \(F^{PE} : \mathcal{L}^{N} \rightarrow 2^{A}\) whose value for a given profile of preferences \(F^{PE}(L_{1}, \ldots, L_{i}, \ldots, L_{N})\) consists of all Pareto efficient alternatives \(a\) for that profile, i.e., \(a \in F^{PE}(L_{1}, \ldots, L_{i}, \ldots, L_{N})\) iff there does not exist \(b \in A\) such that \(b \geq_{L_{i}} a\) for each person \(i\), with strict preference for at least one person \(i\). This is the standard notion of Pareto efficiency and not the idea of unanimity used by Reny and used below. Is \(F^{PE}\) monotonic? We prove that it is by contradiction. Suppose not, i.e., for some \((L_{1}, \ldots, L_{i}, \ldots, L_{N})\) and \(a \in F^{PE}(L_{1}, \ldots, L_{i}, \ldots, L_{N})\), there exists \((L'_{1}, \ldots, L'_{i}, \ldots, L'_{N})\) such that \(a\) does not fall in any person’s ranking in moving from \((L_{1}, \ldots, L_{i}, \ldots, L_{N})\) to \((L'_{1}, \ldots, L'_{i}, \ldots, L'_{N})\), but \(a \notin F^{PE}(L'_{1}, \ldots, L'_{i}, \ldots, L'_{N})\). This means that there exists \(b \in A\) that Pareto dominates \(a\) given the preference profile \((L'_{1}, \ldots, L'_{i}, \ldots, L'_{N})\), i.e.,

\[
\begin{align*}
b & \geq_{L'_{i}} a \\
r \text{for each person } i \text{ with strict inequality for at least one person } i.
\end{align*}
\]

Suppose that \(a >_{L_{i}} b\) for some person \(i\); then

\[
\begin{align*}
a & >_{L_{i}} b
\end{align*}
\]
because $\alpha$ doesn’t fall in any person’s ranking in moving from $(L_1, \ldots, L_i, \ldots, L_N)$ to $(L_1', \ldots, L_i', \ldots, L_N')$. This contradicts the above statement, and so

$$b \geq_{L_i} \alpha$$

for each agent $i$.

To obtain a contradiction to the Pareto efficiency of $\alpha$ given $(L_1, \ldots, L_i, \ldots, L_N)$, we need to show that this last ranking is strict for at least one person $i$. Suppose that

$$b >_{L_i} \alpha$$

but $b =_{L_i} \alpha$. Then $\alpha$ falls in this agent’s ranking relative to $b$ in moving from $(L_1, \ldots, L_i, \ldots, L_N)$ to $(L_1', \ldots, L_i', \ldots, L_N')$, contradicting the hypothesis.