

The paper presents the proof of a result that directly implies both Arrow’s Theorem and the Gibbard-Satterthwaite Theorem. It provides us with the means to discuss these two results and their significance in economic thought. Both theorems have been amply studied and so this topic most likely presents few opportunities for future research. It is worthwhile to understand the implications of these results, however, and the concepts on which the proofs are based may prove useful to you.

**Model**

- **A** - a finite set of alternatives for a group of \( N \in \mathbb{N} \) agents
- **\( L \)** - the set of strict linear orders (or rankings) of elements of **A**
- **\( L^* \)** - the set of weak linear orders of elements of **A**
- an element \((L_1, \ldots, L_i, \ldots, L_N)\) of \( L^N \) is called a profile of rankings
- \( f : L^N \rightarrow A \) - a social choice function
- \( F : L^N \rightarrow L^* \) - a social welfare function

Notice that the use of \( L \) assumes that preferences are unrestricted, i.e., all possible orderings are worthy of consideration.

A social choice function \( f \) represents the process of aggregating individual preferences into a group choice, while a social welfare function \( F \) represents the process of devising a ranking for the group based upon the preferences of the individuals. We might imagine \( f \) as the result of some voting procedure under some solution concept that determines how a person votes based upon his preferences. Obviously, social choice functions exist in reality; we want to explore what properties they should have and what properties they can and do have. A social welfare function fundamentally addresses the meaningfulness of group preferences (e.g., the preferences of a firm or society, or an academic department for that matter). Macroeconomics commonly uses the "representative agent" to represent a large number of people. Is this abstraction meaningful, i.e., is there a sense in which the preferences of the group can be represented as the preferences of an individual?

**Properties of a social choice function:**

- **Unanimity:** If \( a \in A \) is at the top of every individual \( i \)'s ranking \( L_i \), then \( f(L_1, \ldots, L_i, \ldots, L_N) = a \).
- **Dictatorial:** \( f \) is dictatorial if there exists some agent \( i \) such that \( f(L_1, \ldots, L_i, \ldots, L_N) = a \) if and only if \( a \) is the top choice relative to \( L_i \).

This is not a desired property, but it will be part of the statement of Gibbard-Satterthwaite Theorem below.

- **Strategy-Proof:** \( f \) is strategy-proof if for every individual \( i \), every \((L_1, \ldots, L_i, \ldots, L_N)\), and every \( L'_i \),

\[
\begin{align*}
\text{if } f(L_1, \ldots, L_i, \ldots, L_N) \\
\text{is equal to or ranked below} \\
\text{when } i \text{'s ranking is } L_i \text{,}
\end{align*}
\]

\[
\begin{align*}
f(L_1, \ldots, L_i, \ldots, L_N) \\
\text{when } i \text{'s ranking is } L_i \text{,}
\end{align*}
\]

If we regard \( L_i \) as \( i \)'s input into the social choice function that determines an alternative, "strategy-proof" means that \( i \) cannot profit by misreporting and submitting \( L'_i \) rather than his true ranking \( L_i \). This must hold regardless of the reports

\[
(L_1, \ldots, L_{i-1}, L_{i+1} \ldots, L_N)
\]

of the other agent. Honest reporting of one’s ranking is thus a dominant strategy for each agent.
By invoking the \textit{revelation principle} (which will be discussed below), we can see that strategy-proofness is necessary and sufficient for dominant strategy implementation of a social choice function.

**Question:** Is a dictatorship strategy-proof?

**Example 79 (The Borda Count)** Suppose that there are 3 individuals \{1, 2, 3\} and 3 possible alternatives \{a, b, c\}. As in our model, each individual strictly ranks the 3 alternatives. Suppose that the social choice function is determined by asking each individual to assign the numbers 1, 2, and 3 to the 3 alternatives. These are the "votes" of the 3 individuals. The alternative that receives the largest number of votes is the social choice. In the case of ties, the ordering \(a > b > c\) selects among those tied at the top spot. Show that this social function is not strategy-proof.

Consider the rankings
1: \(a > b > c\)
2: \(b > c > a\)
3: \(c > a > b\)

If the individuals vote truthfully, then each candidate gets 3 votes, resulting in the election of \(a\). This is 2's worst choice; if he instead voted \(c > b > a\), then \(c\) would be elected, which he would prefer. The Borda count is therefore not strategy proof.

**Back to the Gibbard-Satterthwaite Theorem**

**Theorem 80 (Gibbard-Satterthwaite Theorem)** If \(A\) has at least 3 elements and \(f : \mathcal{L}^N \rightarrow A\) is onto and strategy-proof, then \(f\) is dictatorial.

**Corollary 81** If \(A\) has at least 3 elements and \(f : \mathcal{L}^N \rightarrow A\) is Pareto optimal and is implemented by a dominant strategy equilibrium of some game \((\prod S_i, \tau)\), then \(f\) is dictatorial.

- **Unanimity and unrestricted preferences ⇒ onto**

The Gibbard-Satterthwaite Theorem is a discouraging result in the sense that it sharply limits what can be accomplished in a strategy-proof manner. The theorem reveals how little can be accomplished through dominant strategy implementation: in a general sense, dominant strategies are too much to ask for, in the sense there very few social choice functions can be implemented in this sense.

History of the result – dates to the 1970’s, discovered independently by Gibbard (Michigan) and Satterthwaite (then a grad student at WI, now at Northwestern). They proved the theorem in different ways, Gibbard deriving it from Arrow’s Theorem and Satterthwaite proving it directly in an exhaustive analysis of cases of preferences. Gibbard published first. Having read Gibbard’s paper, Satterthwaite included a proof in his JET paper that G-S can be used to prove Arrow’s Theorem, thus providing an alternative proof of this result. Gibbard’s result and this result of Satterthwaite together proved the formal equivalence of Arrow’s Theorem and the Gibbard-Satterthwaite Theorem.

Most of the criticism of this result focuses on the assumption that preferences are unrestricted. This applies in many voting problems but not necessarily in economic problems (e.g., "more is better").

**Arrow’s Theorem**

Arrow’s Theorem concerns social welfare functions, not social choice functions, and so we’ll first need to discuss some properties of social welfare functions.

- **Unanimity:** If \(a\) is ranked above \(b\) according to the ranking \(L_i\) for each agent \(i\), then \(a\) is ranked above \(b\) in the ranking \(F(L_1, \ldots, L_n)\).

- **Independence of Irrelevant Alternatives (IIA):** If, for each agent \(i\), the relative ranking of \(a\) and \(b\) is unchanged as \(i\)'s ranking changes from \(L_i\) to \(L'_i\), then the relative ranking of \(a\) and \(b\) is the same in \(F(L_1, \ldots, L_n)\) and \(F(L'_1, \ldots, L'_n)\).

The "irrelevant alternatives" are the other alternatives besides \(a\) and \(b\). Notice that it concerns relative ranking (e.g., \(a\) above \(b\) or \(b\) above \(a\)) and not exact place in the ranking of all \(N\) alternatives.

- **Dictatorial:** \(F\) is dictatorial if there exists some agent \(i\) such that \(F(L_1, \ldots, L_i, \ldots, L_N) = L_i\).

Notice that a dictatorship satisfies independence of irrelevant alternatives. Clearly, we hope for more in a social welfare function than dictatorship.

**Theorem 82** If there are at least 3 alternatives and \(F : \mathcal{L}^N \rightarrow \mathcal{L}^*\) satisfies unanimity and IIA, then \(F\) is dictatorial.
The culprit here is clearly IIA. Most of the subsequent work on Arrow’s Theorem has focused on this hypothesis.

**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 $S_i$, $\sigma_i : L \rightarrow S_i$ 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

$$\left( \prod_{i=1}^{N} S_i, \tau \right).$$

Letting $\sigma : L^N \rightarrow \prod_{i=1}^{N} S_i$ 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 $S_i = L$ (i.e., each agent is given the opportunity to report his complete ranking of the alternatives in $A$).

**Theorem 83 (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 $(L^N, \tau \circ \sigma)$.

**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 $(\prod S_i, \tau)$. 

60
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 84 (Gibbard-Satterthwaite Theorem)** If \( A \) has at least 3 elements and \( f : \mathcal{L}^N \rightarrow A \) is onto and strategy-proof, then \( f \) is dictatorial.

**Corollary 85** If \( A \) has at least 3 elements and \( f : \mathcal{L}^N \rightarrow A \) is Pareto optimal and is implemented by a dominant strategy equilibrium of some game \((\prod S_i, \tau)\), then \( f \) is dictatorial.

**Example 86** A problem from the 2013 midterm, with an additional question to illustrate the revelation principle for Bayesian-Nash equilibrium:

Consider the following two-player game of incomplete information:

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>( \frac{1}{2} )</td>
<td>( L )</td>
<td>( R )</td>
</tr>
<tr>
<td>( T )</td>
<td>( 2\theta_1, 3\theta_2 )</td>
<td>( 1, 1 )</td>
</tr>
<tr>
<td>( B )</td>
<td>( 1, 0 )</td>
<td>( 0, 1 )</td>
</tr>
</tbody>
</table>

It is common knowledge among the two players that each player \( i \)'s type \( \theta_i \) is independently drawn from the uniform distribution on \([0, 1]\).

- Derive a pure strategy Bayesian-Nash equilibrium in this game.
- Illustrate the revelation principle by defining a revelation game that results in the outcome derived for your answer in a) and has the property that truthful revelation is a Bayesian-Nash equilibrium.

We first note that player 1 has a dominant strategy to choose \( T \) for each value of his type \( \theta_1 > 1/2 \) and player 2 has a dominant strategy to choose \( R \) for each value of his type \( \theta_2 < 1/3 \). We therefore conjecture the following form of equilibrium strategies:

\[
\begin{align*}
\text{player 1:} & \quad \left\{ \begin{array}{ll}
T & \text{if } \theta_1 > \theta_1^* \\
B & \text{if } \theta_1 \leq \theta_1^*
\end{array} \right. \\
\text{player 2:} & \quad \left\{ \begin{array}{ll}
L & \text{if } \theta_2 \geq \theta_2^* \\
R & \text{if } \theta_2 < \theta_2^*
\end{array} \right.
\end{align*}
\]

Solving for equilibrium requires solving for the constants \( \theta_1^* \) and \( \theta_2^* \). These are the "critical" values of each player’s type at which he is indifferent between each of his two pure strategies. Notice that player 1 plays \( T \) with probability \( 1 - \theta_1^* \) and \( B \) with probability \( \theta_1^* \), while player 2 plays \( L \) with probability \( 1 - \theta_2^* \) and \( R \) with probability \( \theta_2^* \). We have the equation

\[
2\theta_1^* (1 - \theta_2^*) + 1 \cdot \theta_2^* = 1 \cdot (1 - \theta_2^*) + 0 \cdot \theta_2^*
\]

for player 1, expressing his indifference between \( B \) and \( T \) when \( \theta_1^* \) is his type, and the equation

\[
3\theta_2^* (1 - \theta_1^*) + 0 \cdot \theta_1^* = 1 \cdot (1 - \theta_1^*) + 1 \cdot \theta_1^*
\]

for player 2, expressing his indifference between \( L \) and \( R \) when \( \theta_2^* \) is his type. Player 1’s equation reduces to

\[
\theta_1^* (1 - \theta_2^*) + \theta_2^* = \frac{1}{2} \iff \theta_2^* (1 - \theta_1^*) + \theta_1^* = \frac{1}{2}
\]

and player 2’s equation reduces to

\[
\theta_2^* (1 - \theta_1^*) = \frac{1}{3}.
\]

Substituting the second equation into the first produces

\[
\frac{1}{3} + \theta_1^* = \frac{1}{2} \Rightarrow \theta_1^* = \frac{1}{6}.
\]
We can then solve for $\theta_2^*$ as
\[ \theta_2^* \left( \frac{5}{6} \right) = \frac{1}{3} \Rightarrow \theta_2^* = \frac{2}{5} \]

Given the strategy for player 2 determined by $\theta_2^* = 2/5$, the expected difference for player 1 between choosing $T$ and choosing $B$ when $\theta_1$ is his type equals
\[ \left[ 2\theta_1 \cdot \frac{3}{5} + 1 \cdot \frac{2}{5} \right] - \left[ 1 \cdot \frac{3}{5} + 0 \cdot \frac{2}{5} \right] \]
\[ = \frac{1}{5} [6\theta_1 + 2 - 3] \]
\[ = \frac{1}{5} [6\theta_1 - 1] , \]
which changes from negative to positive at $\theta_1 = 1/6 = \theta_1^*$. This supports his use of the strategy that we have derived for him. Similarly, given the use of this strategy for player 1, the difference in player 2’s expected payoff between $L$ and $R$ when his type is $\theta_2$ equals
\[ \left[ 3\theta_2 \cdot \frac{5}{6} + 0 \cdot \frac{1}{6} \right] - \left[ 1 \cdot \frac{5}{6} + 1 \cdot \frac{1}{6} \right] \]
\[ = \left[ \theta_2 \cdot \frac{5}{2} - 1 \right] . \]
This changes from negative to positive at $\theta_2^* = 2/5$, which supports player 2’s use of the strategy that we derived above.

Our equilibrium is therefore

player 1:
\[
\text{if } \theta_1 > \frac{1}{6} \Rightarrow T, \text{ if } \theta_1 \leq \frac{1}{6} \Rightarrow B. 
\]

player 2:
\[
\text{if } \theta_2 > \frac{2}{5} \Rightarrow L, \text{ if } \theta_2 \leq \frac{2}{5} \Rightarrow R. 
\]

Revelation game: Let $\theta_1^*$ and $\theta_2^*$ now denote the reports of the two agents. The revelation game is defined as follows
\[
\begin{align*}
\theta_1^* \leq \frac{1}{6}, \theta_2^* < \frac{2}{5} & : B,R (0,1) \\
\theta_1^* > \frac{1}{6}, \theta_2^* < \frac{2}{5} & : T,R (1,1) \\
\theta_1^* \leq \frac{1}{6}, \theta_2^* \geq \frac{2}{5} & : B,L (1,0) \\
\theta_1^* > \frac{1}{6}, \theta_2^* \geq \frac{2}{5} & : T,L (2\theta_1,3\theta_2)
\end{align*}
\]