RESEARCH ARTICLE

A one-shot proof of Arrow's theorem and the Gibbard–Satterthwaite theorem

Ning Neil Yu

Abstract This paper provides a simple and transparent proof of a new social choice impossibility theorem. The Gibbard–Satterthwaite theorem and Arrow's impossibility theorem are straightforward corollaries.

1 Introduction

One of the impossibility theorems introduced by Yu (2013) can help prove both the Gibbard–Satterthwaite theorem (Gibbard 1973; Satterthwaite 1975) and Arrow's impossibility theorem (Arrow 1963) succinctly. In this paper, we offer a direct proof of this theorem, which resembles Yu (2012) and employs the "pivotal voter" technique devised and perfected by Barberá (1980), Geanakoplos (2005), and Reny (2001). We comment on different approaches of establishing classical theorems in the last section.

2 The setup

The following terminology follows Yu (2013) closely.

A set of individuals $\mathcal{N} = \{1, \ldots, N\}$ with $N \geq 2$, each have some preferences over $M \geq 3$ alternatives $A = \{a_1, \ldots, a_M\}$. Throughout the paper, we label individuals

N. N. Yu (✉)
Economics Department, Stanford University, 579 Serra Mall, Stanford, CA 94305-6072, USA
e-mail: ningyu@stanford.edu

with n and number alternatives with i, j, k. Let $\mathcal{P}$ be the set of all possible relations on M that are complete, asymmetric, and transitive, i.e., strict preference relations.[1] A *preference profile* is an ordered list $\vec{\succ} = (\succ_1, \ldots, \succ_N)$ with $\succ_n \in \mathcal{P}$ for every $1 \le n \le N$. All possible preference profiles form the collection $\mathcal{P}^{\mathcal{N}}$.

Definition 1 (*SCF*) A *social choice function* $F: \mathcal{P}^{\mathcal{N}} \to A$ assigns to each $\vec{\succ} \in \mathcal{P}^{\mathcal{N}}$ a choice $F(\vec{\succ}) \in A$.

Definition 2 (*SPF*) A *social preference function* $R: \mathcal{P}^{\mathcal{N}} \to \mathcal{P}$ assigns to each $\vec{\succ} \in \mathcal{P}^{\mathcal{N}}$ a preference relation $R(\vec{\succ}) \in \mathcal{P}$.

Several definitions facilitate communication.

Definition 3 For profile $\vec{\succ}$, a_i *dominates* a_j if $a_i \succ_n a_j$ for every n.

Definition 4 Two profiles $\vec{\succ}$ and $\vec{\succ}'$ *agree on* $\{a_i, a_j\}$ if $a_i \succ_n a_j$ is necessary and sufficient for $a_i \succ'_n a_j$ for every n.

Definition 5 $A' \subset A$ is *at the top* of $\vec{\succ}$ if $a_i \succ_n a_j$ for every n, every $a_i \in A'$, and every $a_j \in A \backslash A'$. If in addition $A' = \{a_i, a_j\}$, we call $\vec{\succ}$ a $\{i, j\}$-*runoff*.

Definition 6 A *runoff generating function* $T_{ij}: \mathcal{P}^{\mathcal{N}} \to \mathcal{P}^{\mathcal{N}}$ brings $A' = \{a_i, a_j\}$ to the top, keeping internal rankings of A' and $A \backslash A'$ intact.

3 The main theorem

Definition 7 (*D*) A SCF F is *dictatorial* if there exists a *social choice dictator n* such that $F(\vec{\succ}) \succ_n a_i$ for every $a_i \ne F(\vec{\succ})$ and every $\vec{\succ}$.

Definition 8 (*O*) A SCF F is *onto* if $F(\mathcal{P}^{\mathcal{N}}) = A$.

Definition 9 (*WP*) A SCF F is *weakly Paretian* if a_i dominating a_j implies $F(\vec{\succ}) \ne a_j$ for every i, j and every $\vec{\succ}$.

Definition 10 (*CM*) A SCF F is *Condorcet monotonic* if whenever $\vec{\succ}$ and $\{i, j\}$-runoff $\vec{\succ}'$ agree on $\{a_i, a_j\}$, $F(\vec{\succ}) = a_i$ implies $F(\vec{\succ}') = a_i$.

This condition requires every social choice to be a *generalized Condorcet winner* in that it has to win any runoff it enters that keeps intact its rankings against the opponent. Losing a $\{i, j\}$-runoff prevents a_i from winning any profile that agrees with it on $\{a_i, a_j\}$, no matter what rule—for example, simple majority, super majority, or weighted majority—determines the runoff outcome. In particular, if $F(\vec{\succ}) = a_i$, $F(T_{ij}(\vec{\succ})) = a_i$ by (CM). We often apply the contrapositive statement.

Lemma 1 (O) *and* (CM) *imply* (WP).

Proof Let a_i dominate a_j in $\vec{\succ}$. (O) ensures that $F(\vec{\succ}') = a_i$ for some $\vec{\succ}'$. By (CM), the choice remains a_i for $T_{ik}(\vec{\succ}')$, where $k \notin \{i, j\}$, and further for $T_{ij}(T_{ik}(\vec{\succ}'))$. Since

---

a_j loses the $\{i, j\}$-runoff $T_{ij}(T_{ik}(\vec{\succ}'))$ that agrees with $\vec{\succ}$ on $\{a_i, a_j\}$, the choice for $\vec{\succ}$ cannot be a_j.   □

Theorem 1 *If a SCF F satisfies* (O) *and* (CM)*, then it is dictatorial.*

Proof Pick any $\vec{\succ}'$ with $\{a_i\}$ and $\{a_i, a_j\}$ at the top. Swap the positions of $\{a_i, a_j\}$ sequentially from 1 to N. By (WP), the choice is either a_i or a_j, starting with a_i and ending with a_j. The (i, j)-*pivotal voter* n_{ij} is the first whose swap makes a difference. By (CM), any two (i, j)-runoffs that agree on $\{a_i, a_j\}$ select the same alternative, so n_{ij} is independent of which $\vec{\succ}'$ (with $\{a_i\}$ and $\{a_i, a_j\}$ at the top) we start with.

	1	...	$n_{ij} - 1$	n_{ij}	$n_{ij} + 1$	...	N
$\vec{\succ}''$	a_j	...	a_j	a_i	a_i	...	a_i
	□	...	□				
	a_i	...	a_i	a_j	a_j	...	a_j
	□	...	□	a_k	a_k	...	a_k
$\vec{\succ}'''$	□	...	□				
	a_j	...	a_j	a_j	a_i	...	a_i
	□	...	□		□	...	□
	a_i	...	a_i	a_i	a_j	...	a_j
	□	...	□	a_k	□	...	□

Consider depicted profiles $\vec{\succ}''$ and $\vec{\succ}'''$ with $\{a_i, a_j, a_k\}$ at the top, where columns correspond to voters and squares mark possible positions of a_k. The definition of n_{ij} informs us that $F(T_{ij}(\vec{\succ}'')) = a_i$ and $F(T_{ij}(\vec{\succ}''')) = a_j$, so $F(\vec{\succ}'') \neq a_j$ and $F(\vec{\succ}''') \neq a_i$ by (CM). For $\vec{\succ}''$, the choice is a_i, for (WP) rules out a_k and others. Hence, $F(T_{ik}(\vec{\succ}'')) = a_i$, implying that in defining n_{ik}, no swap before n_{ij} makes a difference, i.e., $n_{ik} \geq n_{ij}$. But j and k are arbitrary, so $n_{ij} = n_{ik}$, i.e., n_{i-} refers to the same individual.

Moreover, $T_{ik}(\vec{\succ}'')$ and $\vec{\succ}'''$ agree on $\{a_i, a_k\}$, so $F(\vec{\succ}''')$ is not a_k due to its loss to a_i in $\{i, k\}$-runoff $T_{ik}(\vec{\succ}'')$. We are left with $F(\vec{\succ}''') = a_j$, so $F(T_{jk}(\vec{\succ}''')) = a_j$, demanding by (CM) that

$$a_j \succ_{n_{ij}} a_k \text{ implies } F(\vec{\succ}) \neq a_k. \tag{*}$$

In defining n_{jk}, (∗) says that no swap before n_{ij} makes a difference, so $n_{jk} \geq n_{ij}$ or $n_{j-} \geq n_{i-}$. But i and j are arbitrary, confirming $n_{j-} = n_{i-}$. The single pivotal voter can eliminate any alternative except her favorite by (∗).   □

4 Classical impossibility theorems

Given Theorem 1, we can prove the Gibbard–Satterthwaite theorem through a simple lemma. Let $(\succ'_n, \vec{\succ}_{-n})$ represent the new profile constructed by replacing the preferences of individual n in $\vec{\succ}$ with $\succ'_n \in \mathcal{P}$.

Definition 11 (*SP*) A SCF F is *strategy-proof* if $F(\succ'_n, \vec{\succsim}_{-n}) \neq F(\vec{\succsim})$ implies $F(\vec{\succsim}) \succ_n F(\succ'_n, \vec{\succsim}_{-n})$ for every n, every $\vec{\succsim}$, and every $\succ'_n$.

This condition demands that an individual is worse off misreporting her preferences whenever a misreport can influence the choice, so truth-telling is optimal.

Lemma 2 (SP) *imply* (CM).

Proof Let $\vec{\succsim}$ and $\{i, j\}$-runoff $\vec{\succsim}'$ agree on $\{a_i, a_j\}$. Suppose that $F(\vec{\succsim}) = a_i$ and $F(\succ'_1, \vec{\succsim}_{-1}) = a_k$ for $k \neq i$. By (SP), $a_i \succ_1 a_k$. Define $\vec{\succsim}'' = (\succ'_1, \vec{\succsim}_{-1})$. Given $\vec{\succsim} = (\succ_1, \vec{\succsim}''_{-1})$, (SP) entails $a_k \succ'_1 a_i$. Since $\vec{\succsim}'$ is a $\{i, j\}$-runoff, $k = j$, contradicting the agreement of $\vec{\succsim}$ and $\vec{\succsim}'$ on $\{a_i, a_j\}$. So $F(\succ'_1, \vec{\succsim}_{-1})$ has to remain a_i. Likewise, the process of updating $\vec{\succsim}$ to $\vec{\succsim}'$ one by one keeps the choice unaltered. □

Theorem 2 (Gibbard–Satterthwaite) *If a SCF F satisfies* (O) *and* (SP)*, then it is dictatorial.*

Proving Arrow's theorem is equally simple.

Definition 12 (*AD*) A SPF R is *Arrow dictatorial* if there exists a *social preference dictator n* such that $a_i \succ_n a_j$ implies $a_i R(\vec{\succsim})a_j$ for every i, j and every $\vec{\succsim}$.

Definition 13 (*AU*) A SPF R is *Arrow unanimous* if $a_i \succ_n a_j$ for every n implies $a_i R(\vec{\succsim})a_j$ for every i, j and every $\vec{\succsim}$.

Definition 14 (*AIIA*) A SPF R is *Arrow independent of irrelevant alternatives* if whenever $\vec{\succsim}$ and $\vec{\succsim}'$ agree on $\{a_i, a_j\}$, $a_i R(\vec{\succsim})a_j$ implies $a_i R(\vec{\succsim}')a_j$.

Theorem 3 (Arrow's) *If a SPF R satisfies* (AU) *and* (AIIA)*, then it is Arrow dictatorial.*

Proof Given R, we can define a SCF F^R that selects the alternative ranked highest by R. F^R obviously satisfies (O), because by (AU), $F^R(\vec{\succsim}) = a_i$ if only $\{a_i\}$ is at the top. To see (CM), let $\vec{\succsim}$ and $\{i, j\}$-runoff $\vec{\succsim}'$ agree on $\{a_i, a_j\}$. When $F^R(\vec{\succsim}) = a_i$, $a_i R(\vec{\succsim})a_j$, so by (AIIA), $a_i R(\vec{\succsim}')a_j$. But $a_i R(\vec{\succsim}')a_k$ for $k \notin \{i, j\}$ by (AU), so $F^R(\vec{\succsim}')$ has to be a_i.

Theorem 1 thus presents a social choice dictator n of F^R. She is a social preference dictator too. If $a_i \succ_n a_j$, individual n dictates $F^R(T_{ij}(\vec{\succsim})) = a_i$, so $a_i R(T_{ij}(\vec{\succsim}))a_j$ and by (AIIA) $a_i R(\vec{\succsim})a_j$. □

5 Concluding remarks

The simplicity is no coincidence. Our main theorem employs assumptions that are almost logically minimal (Yu 2013). Traditionally, people prove the Gibbard–Satterthwaite theorem as a corollary of the Muller–Satterthwaite theorem (Muller and Satterthwaite 1977). But the "monotonicity" condition assumed is more restrictive than (CM), making it a weaker theorem than ours. Reny (2001) has to circumvent the difficulty by giving parallel proofs of Arrow's theorem and the Muller–Satterthwaite

theorem. Mas-Colell et al. (1995) prove Arrow's theorem first, but as a step toward the Muller–Satterthwaite theorem, the construction of R^F, a SPF derived from a SCF, is necessarily more involved than that of F^R. Building on these earlier works,[2] our framework offers an alternative way of unifying and teaching classical theorems.

References

Arrow, K.J.: Social Choice and Individual Values, 2nd edn. Wiley, New York (1963)
Barberá, S.: Pivotal voters: a new proof of Arrow's theorem. Econ. Lett. **6**(1), 13–16 (1980)
Geanakoplos, J.: Three brief proofs of Arrow's impossibility theorem. Econ. Theory **26**(1), 211–215 (2005)
Gibbard, A.: Manipulation of voting schemes: a general result. Econometrica **41**(4), 587–601 (1973)
Man, P., Takayama, S.: A unifying impossibility theorem. Econ. Theory (2013, in press). doi:10.1007/s00199-013-0750-6
Mas-Colell, A., Whinston, M., Green, J.: Microeconomic Theory. Oxford University Press, Oxford (1995)
Muller, E., Satterthwaite, M.A.: The equivalence of strong positive association and strategy-proofness. J. Econ. Theory **14**(2), 412–418 (1977)
Reny, P.J.: Arrow's theorem and the Gibbard–Satterthwaite theorem: a unified approach. Econ. Lett. **70**(1), 99–105 (2001)
Satterthwaite, M.A.: Strategy-proofness and Arrow's conditions: existence and correspondence theorems for voting procedures and social welfare functions. J. Econ. Theory **10**(2), 187–217 (1975)
Yu, N.N.: A one-shot proof of Arrow's impossibility theorem. Econ. Theory **50**(2), 523–525 (2012)
Yu, N.N.: A quest for fundamental theorems of social choice. Working Paper (2013)

---

[2] Another source of our inspiration is Man and Takayama (2013). They also derive other results from one main theorem, the proof of which is relatively long, partly because it works with neither SCFs nor SPFs.