THREE BRIEF PROOFS OF ARROW'S IMPOSSIBILITY THEOREM

John Geanakoplos*

Revised: August 1996

ABSTRACT

Arrow's original proof of his impossibility theorem proceeded in two steps: showing the existence of a decisive voter, and then showing that a decisive voter is a dictator. Barbera replaced the decisive voter with the weaker notion of a pivotal voter, thereby shortening the first step, but complicating the second step. I give three brief proofs, all of which turn on replacing the decisive/pivotal voter with an extremely pivotal voter (a voter who by unilaterally changing his vote can move some alternative from the bottom of the social ranking to the top), thereby simplifying both steps in Arrow's proof. My first proof uses almost no notation, while the second uses May's notation and is extremely brief. The third proof is perhaps the most interesting, because along the way to proving the existence of an extremely pivotal voter, it shows that the Arrow axioms guarantee issue neutrality, that is, that every choice must be made by exactly the same process.

*I wish to thank Ken Arrow, Don Brown, Ben Polak, Herb Scarf, Chris Shannon, Lin Zhou, and especially Eric Maskin for very helpful comments and advice. I was motivated to think of reproving Arrow's theorem when I undertook to teach it to George Zettler, a mathematician friend. After I presented this paper at MIT, a graduate student there named Luis Ubeda–Rives told me he had worked out the same neutrality argument as I give in my third proof while he was in Spain nine years ago. He said he was anxious to publish on his own and not jointly, so I encourage the reader to consult his forthcoming working paper.

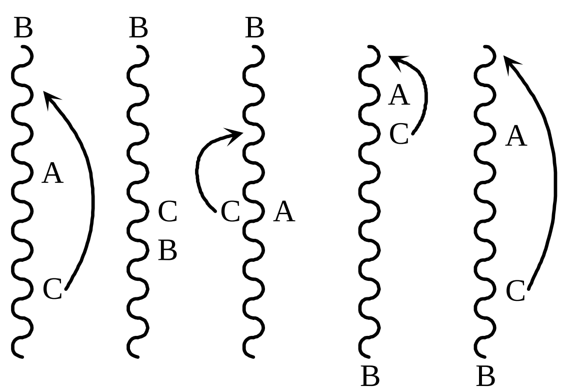
1 Informal Statement and Proof of Arrow's Theorem

Let $\mathbf{A} = \{A, B, ..., C\}$ be a finite set of at least three alternatives. A transitive preference over $\mathbf{A}$ is a ranking of the alternatives in $\mathbf{A}$ from top to bottom, with ties allowed. We consider a society with N individuals, each of whom has a (potentially different) transitive preference. A constitution is a function which associates with every N-tuple (or profile) of transitive preferences a transitive preference called the social preference.

A constitution respects **unanimity** if society puts alternative α above β whenever every individual puts α above β. The constitution respects **independence of irrelevant alternatives** if the social relative ranking (higher, lower, or indifferent) of two alternatives α and β depends only on their relative ranking by every individual. The constitution is a **dictatorship** by individual n if society strictly prefers α to β whenever n strictly prefers α to β.

ARROW'S THEOREM: *Any constitution that respects transitivity, independence of irrelevant alternatives, and unanimity is a dictatorship.*

FIRST PROOF: Let alternative B be chosen arbitrarily. We argue first that at any profile in which every voter puts alternative B at the top or bottom of his ranking of alternatives, society must as well (even if half the voters put B at the top and half put B at the bottom). Suppose to the contrary that for such a profile and for distinct A, B, C, the social preference put $A \geq B$ and $B \geq C$. By independence of irrelevant alternatives, this would continue to hold even if every individual moved C above A, because that could be arranged without disturbing any AB or CB votes (since B occupies an extreme position in each individual's ranking, as can be seen from the diagram). By transitivity the social ranking would then put $A \geq C$, but by unanimity it would also put $C > A$, a contradiction.



We argue that there is a voter $n^* = n(B)$ who is extremely pivotal in the sense that by changing his vote at some profile he can move B from the bottom of the social ranking to the top. To see this, let each voter put B at the bottom of his (otherwise arbitrary) ranking of alternatives. By unanimity, society must as well. Now let the individuals from voter 1 to N successively move B from the bottom of their rankings to the top, leaving the other relative rankings in place. Let n^* be the first voter whose change causes the social ranking of B to change. (By unanimity, a change must occur at the latest when $n^* = N$.) Denoted by profile I the list of all voter rankings just before n^* moves B, and denote by profile II the list of all voter rankings just after n^* moves B to the top. Since in profile II B has moved off the bottom of the social ranking, we deduce from our first argument that the social preference corresponding to profile II must put B at the top.

We argue third that $n^* = n(B)$ is a dictator over any pair AC not involving B. To see this, choose one element, say A, from the pair AC. Construct profile III from profile II by letting n^* move A above B, so that $A >_{n^*}> B >_{n^*} C$, and by letting all the agents $n \neq n^*$ arbitrarily rearrange their relative rankings of A and C while leaving B in its extreme position. By independence of irrelevant alternatives, the social preferences corresponding to profile III would necessarily put $A > B$ (since all individual AB votes are as in profile I where n^* put B at the bottom), and $B > C$ (since all individual BC votes are as in profile II where n^* put B at the top). By transitivity, society must put $A > C$. By independence of irrelevant alternatives, the social preference over AC must agree with n^* whenever $A >_{n^*} C$.

We conclude by arguing that n^* is also a dictator over every pair AB. Given any alternative A different from B, we can find a third distinct alternative C to put at the bottom in the above construction. From the above argument, there must be a voter $n(C)$ who is an $\alpha\beta$ dictator for any pair $\alpha\beta$ not involving C, such as AB. But agent n^* can affect society's AB ranking, namely at profiles I and II, hence this AB dictator $n(C)$ must actually be n^*. $\square$

2 Arrow's Theorem Formally Stated

Let $\mathcal{A} = \{A, B, C\}$ be a set of three alternatives.[1] A **transitive** preference π is a binary relation that specifies for each pair of alternatives $\alpha\beta \in \mathcal{A} \times \mathcal{A}$ whether the first is preferred to the second (which we write as $\pi_{\alpha\beta} = 1$) or the second is preferred to the first (which we write as $\pi_{\alpha\beta} = 0$) or the two are indifferent (which we write as $\pi_{\alpha\beta} = \frac{1}{2}$), and that has no cycles. Since $\pi_{\alpha\beta} = 1 - \pi_{\beta\alpha}$, a preference π can be identified with a triple $(\pi_{AB}, \pi_{BC}, \pi_{CA}) \in \{0, \frac{1}{2}, 1\} \times \{0, \frac{1}{2}, 1\} \times \{0, \frac{1}{2}, 1\}$. For example, (110) means $A\pi B$, $B\pi C$, and $A\pi C$, and corresponds to the ranking $\begin{pmatrix} A \\ B \\ C \end{pmatrix}$. The set of transitive preferences is $\Pi = \{abc \in \{0,1\}^3 \,:\, abc \neq 000 \text{ and } abc \neq 111\} \cup \left\{abc \in \{0, \frac{1}{2}, 1\}^3 \,:\, a + b + c = \frac{3}{2}\right\} = \left\{(abc) \in \{0, \frac{1}{2}, 1\}^3 \,:\, abc = \frac{1}{2}\,\frac{1}{2}\,\frac{1}{2} \text{ or } \{0, 1\} \subset \{a, b, c\}\right\}$.

[1]For the purposes of Arrow's Impossibility Theorem, there is no loss of generality in restricting attention to three alternatives. If every subset of three alternatives is determined by a dictator, then because such sets overlap, Arrow's axiom of independence of irrelevant alternatives guarantees that one dictator would determine all social preferences.

The set of agents $n \in N$ is a finite set. We call Π^N the set of legal preferences.

A **constitution** is a function $(f_1,\, f_2,\, f_3) = f\,:\,\Pi^N \to \Pi$, which we denote

$$f(\boldsymbol{\pi}) = f\begin{pmatrix}\boldsymbol{\pi}(1)\\ \vdots \\ \boldsymbol{\pi}(n)\\ \vdots \\ \boldsymbol{\pi}(N)\end{pmatrix} = f\begin{bmatrix}\boldsymbol{\pi}_{AB}(1) & \boldsymbol{\pi}_{BC}(1) & \boldsymbol{\pi}_{CA}(1)\\ \vdots & & \\ \boldsymbol{\pi}_{AB}(n) & \boldsymbol{\pi}_{BC}(n) & \boldsymbol{\pi}_{CA}(n)\\ \vdots & & \\ \boldsymbol{\pi}_{AB}(N) & \boldsymbol{\pi}_{BC}(N) & \boldsymbol{\pi}_{CA}(N)\end{bmatrix} = f\begin{bmatrix}\boldsymbol{\pi}_1(1) & \boldsymbol{\pi}_2(1) & \boldsymbol{\pi}_3(1)\\ \vdots & & \\ \boldsymbol{\pi}_1(n) & \boldsymbol{\pi}_2(n) & \boldsymbol{\pi}_3(n)\\ \vdots & & \\ \boldsymbol{\pi}_1(N) & \boldsymbol{\pi}_2(N) & \boldsymbol{\pi}_3(N)\end{bmatrix} = f[\boldsymbol{\pi}_1\,\boldsymbol{\pi}_2\,\boldsymbol{\pi}_3].$$

The rows of the matrix $\boldsymbol{\pi}$ represent individual preferences, and the columns represent all individual preferences over one pair of alternatives.

The constitution f satisfies **independence of irrelevant alternatives** (IIA) if for each $i \in \{AB,\, BC,\, CA\}$, f_i depends only on the ith column of $\boldsymbol{\pi}$: $f_i(\boldsymbol{\pi}) = f_i(\boldsymbol{\omega})$ whenever $\boldsymbol{\pi}_i(n) = \boldsymbol{\omega}_i(n)$ for all $n \in N$. We will often abuse notation and write $f_i(\boldsymbol{\pi}_i)$ or $f_i(\mu)$ where $\boldsymbol{\pi}_i = \mu \in \{0, \frac{1}{2}, 1\}^N$ is just the ith column of $\boldsymbol{\pi}$.

Let 1 denote the vector of all 1's, and 0 denote the vector of all 0's. (The context will distinguish between 1 as vector and scalar.) **Unanimity** (U) requires $f_i(1) = 1$ and $f_i(0) = 0$ for $i = 1,\, 2,\, 3$.

If $\mu \in \{0, \frac{1}{2}, 1\}^N$, then $\mu(-n) \in \{0,\, \frac{1}{2},\, 1\}^{N-1}$ denotes the vector obtained by deleting $\mu(n)$ from μ, and $\binom{a}{\mu(-n)} \in \{0, \frac{1}{2}, 1\}^N$ denotes the vector obtained from μ by changing $\mu(n)$ to a. We say that a constitution f makes an agent $n \in N$ pivotal in coordinate i at a particular $\mu(-n) \in \{0,\, \frac{1}{2},\, 1\}^{N-1}$ if $f_i\binom{1}{\mu(-n)} = 1$ and $f_i\binom{0}{\mu(-n)} = 0$. We say that f makes n a dictator iff f makes n pivotal in every coordinate $i \in \{1,\, 2,\, 3\}$ at every $\mu(-n) \in \{0,\, \frac{1}{2},\, 1\}^{N-1}$.

ARROW'S IMPOSSIBILITY THEOREM: *Let $f\,:\,\Pi^N \to \Pi$ be a constitution satisfying unanimity and independence of irrelevant alternatives, where Π is the set of all transitive preferences on a set of three or more alternatives. Then f makes some agent $n \in N$ a dictator.*

3 Second Proof [2]

LEMMA 1: *Let $\boldsymbol{\pi} \in \Pi^N$ and suppose that f makes an agent $n \in N$ pivotal in i at $\boldsymbol{\pi}_i(-n)$ and pivotal in $j \neq i$ at $\boldsymbol{\pi}_j(-n)$. Then f makes n pivotal in the third coordinate $k \notin \{i, j\}$ at $\boldsymbol{\pi}_k(-n)$.*

PROOF OF LEMMA 1: WLOG, $i = 1$, $j = 2$. Since $f\binom{110}{\boldsymbol{\pi}(-n)} = (11a) \in \Pi$, $a = 0$. Similarly, $\binom{001}{\boldsymbol{\pi}(-n)} = (00b) \in \Pi$, so $b = 1$. By IIA, n is also pivotal in the third coordinate. $\square$

[2] Rather than repeating the first proof with the notation of the last section, we give a variation on the theme. In particular, we use the Condorcet triple to show the existence of an extremely pivotal voter.

LEMMA 2: *Let $\boldsymbol{\pi} \in \Pi^N$ and suppose f makes n pivotal at $\boldsymbol{\pi}_i(-n)$ for all $i = 1, 2, 3$. Then n is a dictator.*

PROOF OF LEMMA 2: Change one entry of one column of $\boldsymbol{\pi}(-n)$ legally. By IIA, n is still pivotal in the other two coordinates, hence by Lemma 1, in all three. By changing one entry at a time, one can move from $\boldsymbol{\pi}(-n)$ to any other $\boldsymbol{\omega}(-n) \in \Pi^{N-1}$ without leaving Π^{N-1}. $\square$

LEMMA 3: *There exists $n \in N$ and $\boldsymbol{\pi} \in \Pi^N$ such that f makes n pivotal in coordinate i at $\boldsymbol{\pi}_i(-n)$ for $i = 1$ and 2.*

PROOF OF LEMMA 3: Let $\Delta \equiv \{\boldsymbol{\pi} \in \Pi^N : f(\boldsymbol{\pi}) = 110$, and for all $n \in N$, $\boldsymbol{\pi}(n) = 110$ or 101 or $011\}$.[3] Note that by unanimity $\begin{bmatrix} 110 \\ 110 \end{bmatrix} \in \Delta$. Let $\boldsymbol{\pi}$ be any element of Δ with the fewest number of agents n with $\boldsymbol{\pi}(n) = 110$. Note that there must be at least one such agent n, for if not, then by unanimity and IIA, $f_3(\boldsymbol{\pi}) = 1$, contradicting $\boldsymbol{\pi} \in \Delta$.

Observe that by construction and IIA, $f\begin{bmatrix} 110 \\ \boldsymbol{\pi}(-n) \end{bmatrix} = 110$, while $f\begin{bmatrix} 011 \\ \boldsymbol{\pi}(-n) \end{bmatrix} = a1c \neq 110$ and $f\begin{bmatrix} 101 \\ \boldsymbol{\pi}(-n) \end{bmatrix} = 1bc \neq 110$. We must show $a = b = 0$. If $c = 0$, then $a \neq 1$ and $b \neq 1$. But then by IIA, $f\begin{bmatrix} 001 \\ \boldsymbol{\pi}(-n) \end{bmatrix} = abc \leq \left(\frac{1}{2}\,\frac{1}{2}\,0\right)$, contradicting transitivity. So $c \geq \frac{1}{2}$. Then $a1c \in \Pi$ implies $a = 0$ and $1bc \in \Pi$ implies $b = 0$. $\square$

4 Third Proof

For any vector $\alpha \in \{0, \frac{1}{2}, 1\}^K$ for $K = N$ or $N-1$ or 3 or 1, we denote by $\bar{\alpha}$ the vector obtained by changing each coordinate α_i into $\bar{\alpha}_i = 1 - \alpha_i$. Thus $\bar{1} = 0$ and $\bar{0} = 1$.

LEMMA 1 (Strict individual preferences give strict and neutral social preferences)**:** *There is a function $g : \{0,1\}^N \to \{0,1\}$ such that for all $i \in \{1,2,3\}$ and all $\mu \in \{0,1\}^N$, $f_i(\mu) = g(\mu) = \bar{g}(\bar{\mu})$. In other words, $f_{AB}(\mu) = f_{BC}(\mu) = f_{CA}(\mu) = f_{BA}(\mu) \equiv \bar{f}_{AB}(\bar{\mu}) = f_{CB}(\mu) \equiv \bar{f}_{BC}(\bar{\mu}) = f_{AC}(\mu) \equiv \bar{f}_{CA}(\bar{\mu}) \neq \frac{1}{2}$.*

PROOF of LEMMA 1: Suppose $f_1(\mu) \geq \frac{1}{2}$ and $f_2(\bar{\mu}) \geq \frac{1}{2}$ for some $\mu \in \{0,1\}^N$. Then $[\mu\bar{\mu}1]$ is legal, since every row contains a 0 and 1, and $f[\mu\bar{\mu}1] \geq (\frac{1}{2}\,\frac{1}{2}\,1)$, contradicting transitivity. We conclude that $\forall \mu \in \{0,1\}^N$, $f_1 = \bar{f}_2(\bar{\mu}) \neq \frac{1}{2}$. Similarly, $f_3(\mu) = \bar{f}_2(\bar{\mu})\ \forall \mu \in \{0,1\}^N$. Since the subscripts were arbitrary, the lemma is proved. $\square$

[3] The reader should recognize 110, 101, 011 as the Condorcet triple. Note that each preference in the Condorcet triple can be obtained from the next preference by moving the bottom alternative to the top.

LEMMA 2: *Some agent is symmetrically doubly pivotal: there is $n \in N$ and $\mu(-n) \in \{0,1\}^{N-1}$ such that $g\binom{1}{\mu(-n)} = g\binom{1}{\bar{\mu}(-n)} = 1$ and $g\binom{0}{\mu(-n)} = g\binom{0}{\bar{\mu}(-n)} = 0$.*

PROOF OF LEMMA 2: By unanimity, $g(1) = 1$ and $g(0) = 0$. Change the coordinates one-by-one from 1 to 0; for some n, Lemma 1 assures us that g must switch all the way to 0. Take $\mu(-n) = (\underbrace{0,...,0}_{1 \text{ to } n-1}, \underbrace{0,...,0}_{n+1 \text{ to } N})'$. Then $g\binom{1}{\mu(-n)} = 1$ and $g\binom{0}{\mu(-n)} = 0$. From the neutrality part of Lemma 1, $g\binom{0}{\bar{\mu}(-n)} = \overline{1} = 0$ and $g\binom{1}{\bar{\mu}(-n)} = \overline{0} = 1$. □

PROOF OF THEOREM: Take any $\eta(-n) \in \{0, \frac{1}{2}, 1\}^{N-1}$; then $\boldsymbol{\pi} \equiv \begin{bmatrix} 1 & 1 & 0 \\ \mu(-n) & \bar{\mu}(-n) & \eta(-n) \end{bmatrix}$ is legal, since $\mu(-n)$ was constructed to lie in $\{0,1\}^{N-1}$, so that every row of $\boldsymbol{\pi}$ contains a 0 and a 1. By Lemma 2, $f(\boldsymbol{\pi}) = 11 f_3\binom{0}{\eta(-n)} \in \Pi$. We conclude that $f_3\binom{0}{\eta(-n)} = 0$. Similarly, $\begin{bmatrix} 0 & 0 & 1 \\ \mu(-n) & \bar{\mu}(-n) & \eta(-n) \end{bmatrix}$ is legal, so $f_3\binom{1}{\eta(-n)} = 1$. Hence n is a dictator in issue 3; similarly, agent n is a dictator in issues 1 and 2. □

References

Arrow, Kenneth. 1951. *Social Choice and Individual Values.* New York: John Wiley and Sons.

Barbera, Salvador, 1980. "Pivotal Voters. A New Proof of Arrow's Theorem," *Economics Letters*, 6, 13–16.

May, K. 1952. "A Set of Independent, Necessary, and Sufficient Conditions for Simple Majority Decisions," *Econometrica*, 20, 680–684.