Proof of Arrow's Impossibility Theorem

From: J. Kelly, *Social Choice Theory: An Introduction*
If you are interested in more detail on axiomatic social choice theory,
Kelly's book is the best place to start.

I. **Arrow's General Impossibility Theorem**

A. **The Axioms**: Properties we would want a social choice function to have

1. **Collective Rationality**: the social preference relation is reflexive, transitive and complete.

a. Kelly (1988) says that a choice function is *explicable* if there is a relation, Ω, such that

$$C(\boldsymbol{v}) = \{x \in \boldsymbol{v}: x \,\Omega\, y, \forall\, y \in \boldsymbol{v}\}$$

b. Furthermore, a choice function has a *transitive explanation* if Ω is reflexive, complete, and transitive.

c. Then we say that a social choice rule (or function) has transitive explanations if at every admissible profile, $\boldsymbol{R}$, the associated $C_{\boldsymbol{R}}$ has a transitive explanation.

d. Essentially, this says that we want our social choice function to have the same minimal property of rationality that individuals have.

(1) Specifically, it says we don't want any cycles of the sort that Condorcet identified.

(2) Note the implication: collective rationality rules out the Condorcet producedure.

2. **Unrestricted Domain**: the social preference function has as its domain all logically possible profiles of preference orderings on $\boldsymbol{O}$ and all possible agendas $\boldsymbol{v} \subset \boldsymbol{O}$.

a. This says that, at least *a priori*, we have no reason in a democratic political system to define any particular individual preference ordering, and thus we cannot rule out any particular profile.

b. Note that this may be violated by real societies where, say, socialization makes certain profiles very unlikely.

3. **The Pareto Property**

a. Variants of the Pareto Property

(1) **Weak Pareto Principle**: For x and $y \in \boldsymbol{O}$, if $x P_i y \ \forall i \in I$, then $x P_S y$.

(2) **Strong Pareto Principle**: Let the social choice rule select choice function $C_{\boldsymbol{R}}$ at profile $\boldsymbol{R}$. Suppose that at $\boldsymbol{R}$ everyone unanimously finds one alternative, x, to be at least as good as y ($x R_i y \ \forall i \in I$), and at least one individual strictly prefers x to y ($x P_i y$ for at least one person). Then, if x is available ($x \in \boldsymbol{v}$), y won't be chosen ($y \notin C_{\boldsymbol{R}}(\boldsymbol{v})$).

(a) The strong Pareto principle is "strong" in the sense that it excludes more alternatives from the chosen set than the weak pareto principle.

b. These insist on a very weak form of democratic-ness. Specifically, these are attempts to implement formally the notion that democratic social choice rules should be *positively responsive* to preferences.

c. The Pareto principle captures notions like

(1) **Monotonicity**: if some individual raises her evaluation of a chosen alternative (all other evaluations held constant), that alternative cannot cease to be chosen; or, if some individual lowers her evaluation of a non-chosen alternative (*ceteris paribus*) that alternative cannot become chosen.

(2) **Non-Imposition**

4. **Independence of Irrelevant Alternatives** (IIA): If $\boldsymbol{R}$ is a profile over some set of alternatives that includes x and y, if $G(\boldsymbol{R}, \{x,y\}) = x P_S y$ (i.e. $C_{\boldsymbol{R}}(\{x,y\}) = x$), and if $\boldsymbol{R}'$ is another preference profile such that each individual's preferences between x and y are unchanged from the first profile, then $G(\boldsymbol{R}', \{x,y\}) = x P_S y$.

a. Example of failure of independence for the Borda rule

(1) Consider the 3 person profile **R**

(a) xyzw

(b) yzwx
(c) zwxy

(d) With 4 alternatives, we assign 4 points for a first-place ranking, 3 for a second, 2 for a third, and 1 for fourth. Thus, the Borda scores for this profile are:

i) w = 6
ii) x = 7
iii) y = 8
iv) z = 9

(2) Here is another 3 person profile, $\boldsymbol{R}'$

(a) xzyw
(b) ywxz
(c) wxzy

(d) The Borda scores here are

i) w = 8
ii) x = 9
iii) y = 7
iv) z = 6

(3) Now consider the agenda $\boldsymbol{v} = \{x,y,w\}$, i.e. delete z.

(a) $G(\boldsymbol{R}.\ \boldsymbol{v}) = y$, and $G(\boldsymbol{R}', \boldsymbol{v}) = x$.

(b) This is problematic since we are getting different choices from $\boldsymbol{v}$ even though the two profiles agree completely over this agenda.

b. The assumption of independence says that this is an unattractive property for a social choice rule so, if two profiles $\boldsymbol{R}$ and $\boldsymbol{R}'$, restricted to an agenda $\boldsymbol{v}$ are identical, then the choices made from that agenda should be the same.

c. Let me just warn against a common misunderstanding of this condition: independence does not rule out "intensity" of preference in making social choices.

(1) It is part of our definition of a social choice rule/function that

the choices are based only on the information in a profile of ordinal preference relations.

(2) These preference relations do not contain any intensity information that could be used by social choice rules, whether or not they violate the independence axiom.

d. Note the *difference between transitivity and independence*

(1) to check whether there is a transitive explanation you fix the profile and vary the agenda; while

(2) to check independence you fix the agenda and vary the profile.

5. **Nondictatorship**: No person *i* is decisive for every pair of outcomes in ***O***.

a. **Decisiveness**: A group *g*, or individual *i*, is said to be decisive for alternate *x* against alternate *y* if at every profile in the domain of the rule, if $x P_i y$ for individual *i*, then for any agenda containing *x*, *y* will not be chosen--even if all non-*g*'s prefer *y* to *x*.

b. If individual *i* is decisive for every pair of alternatives in ***O***, we say that the individual is a dictator.

B. **Theorem** (Arrow, 1950, *JPE*): If ***O*** consists of 3 or more outcomes, the only rules that satisfy collective rationality, unrestricted domain, the pareto principle, and independence of irrelevant alternatives, violate nondictatorship.

1. Think about what this says: it doesn't say that it is difficult to find a social choice rule that satisfies these five axioms, it says it's impossible.

2. This is all the more striking because this is a fairly short list of axioms. Real world social choice rules have many more properties than these.

a. e.g. we might want the rule to select a *unique* alternative; or we might want to grant individuals decisive power over certain classes of decision (e.g. we might want to let people decline elective office).

b. In fact, Arrow set out to prove that such functions **do** exist.

3. Furthermore, all of the axioms attempt to get at things that we would generally take to be desirable properties of social choice rules.

C. **Proof**:

1. The strategy of the proof is to assume that all five conditions hold and derive a contradiction.

a. This will prove that the assumption that the five conditions hold is false.

b. Following Kelly (1988) we will develop three preliminary results called **contagion theorems**.

c. A little jargon and notation:

(1) Suppose that there are N individuals, a subset $T \subset N$ is **locally decisive** for alternative x against alternative y if, whenever profile **R** satisfies

(a) $x\ R_i\ y, \forall\ i \in T$;
(b) $x\ P_i\ y$, for at least one $i \in T$; and
(c) $y\ P_j\ x, \forall\ j \in N - T$;

then $x \in \mathbf{v}$ implies $y \notin C_{\mathbf{R}}(\mathbf{v})$. Note that this is only exclusionary power, not the power to select.

(2) If T is *locally decisive* for x against y, T can exclude y (if x is available and members of T prefer x to y) but this exclusion only takes effect if everyone ouside T strongly prefers y to x.

(3) A set is called *decisive* (or **globally decisive**) for x against y if it can exclude y no matter what pattern of preferences on x and y are held by people not in T.

(4) If a set T is locally decisive for x against y we will write $x\ D_T\ y$; if T is globally decisive against y we will write $x\ D_T^*\ y$.

2. **Lemma 1** (First Narrow Contagion Result): Suppose with at least three individuals and at least three alternatives, a social choice rule satisfies collective rationality, unrestricted domain, strong Pareto condition, and IIA. If for this rule T is locally decisive for a against b, then T is globally decisive for a against c, where a, b, and c are distinct alternatives in **O**.

Proof: Assume $a\ D_T\ b$ and we seek to prove that $a\ D_T^*\ c$.

a. To prove $a\ D_T^*\ c$, we must show $a \in \mathbf{v}$ implies $c \notin \mathbf{v}$ at *any* profile **R**

where everyone in T finds a to be at least as good as c and one strictly prefers a.

b. So let **R** be such a profile and partition T into T_1 and T_2, such that everyone in T_1 has a P_i c and everyone in T_2 has a I_i b. The remaining individuals in N, N-T, may have any orderings of a and c (in particular, they may have the same or different orderings). In Kelly's notation this is

(1) T_1: ac
(2) T_2: (ac)
(3) N-T: [ac]

c. Now we want to show that $a \in v$ implies $c \notin C_{\mathbf{R}}(\mathbf{v})$. Since $C_{\mathbf{R}}$ has some transitive explanation by Ω, it is equivalent to show $a \Omega c$ and not $c \Omega a$.

(1) It can be shown that if unrestricted domain holds, there is a *unique* Ω that explains $C_{\mathbf{R}}$, namely

$x \Omega y$ iff $x \in C_{\mathbf{R}}(\{x,y\})$.

(2) Thus, it is sufficient to prove $C_{\mathbf{R}}(\{a,c)\} = \{a\}$.

(3) To prove this, we consider a new profile, constructed to be very closely related to **R**.

(a) Profile $\mathbf{R}'$, restricted to {a,b,c} is

(b) T_1: abc
(c) T_2: (abc)
(d) N-T: b[ac]

(e) In this case the square brackets imply that, whatever ranking an individual in N-T had of a and c at **R**, it is unchanged at $\mathbf{R}'$ by the addition of b, which dominates both.

(4) By unrestricted domain, the social choiice rule yields a choice function $C_{\mathbf{R}'}$.

(5) Another application of the domain constraint tells us that {a,b}, {a,c} and {b,c} are in the domain of $C_{\mathbf{R}'}$.

(6) By local decisiveness of T for a against b (and noting that

everyone in T finds a R_i b, someone in T_1 finds a P_i b, and everyone in N-T finds b P_i a) we get $C_{\mathbf{R}'}(\{a,b\}) = a$.

(7) From the strong Pareto condition (and noting that at $\mathbf{R}'$ everyone finds b to be at least as good as c--b R_i c $\forall$ i $\in$ I) we get $C_{\mathbf{R}'}(\{b,c\}) = \{b\}$.

d. From $C_{\mathbf{R}'}(\{a,b\}) = \{a\}$ and $C_{\mathbf{R}'}(\{b,c\}) = \{b\}$, we now want to show that $C_{\mathbf{R}'}(\{a,c\}) = \{a\}$.

(1) First, we exploit our knowledge that $C_{\mathbf{R}'}$ has some reflexive, complete and transitive explanation, Ω':

(a) From $C_{\mathbf{R}'}(\{a,b\}) = \{a\}$ we get a Ω' b;
(b) From $C_{\mathbf{R}'}(\{b,c\}) = \{b\}$ we get b Ω' c;
(c) These, with transitivity of Ω' give a Ω' c.
(d) By reflexivity a Ω' a.

(2) Together with the fact that Ω' explains $C_{\mathbf{R}'}$, these tell us that that a $\in C_{\mathbf{R}'}(\{a,c\})$.

(3) To complete the proof, we want to show that c $\notin C_{\mathbf{R}'}(\{a,c\})$.

(a) Suppose not. Then c Ω' a.

(b) Since also b Ω' c, transitivity gives b Ω' a.

(c) Together with b Ω' b, this would tell us that b$\in$ $C_{\mathbf{R}'}(\{a,b\})$, which is false.

(d) Thus c $\notin C_{\mathbf{R}'}(\{a,c\})$ and so $C_{\mathbf{R}'}(\{a,c\}) = \{c\}$.

e. Finally, since **R** and $\mathbf{R}'$ agree on {a,c}, one application of IIA yields

$$C_{\mathbf{R}}(\{a,c\}) = \{a\},$$

which is what we wanted to show■

3. **Lemma 2** (Second Narrow Contagion Result): Suppose with at least three individuals and at least three alternatives a social choice rule satisfies unrestricted domain, strong Pareto condition, IIA, and has transitive explanations. If for this rule a set T is locally decisive for a against b, then T is globally decisive for c against b, where a, b, and c are distinct alternatives in **O**.

Proof: Just like the first narrow contagion result.

4. **Lemma 3** (Broad Contagion Result): Suppose with at least three individuals and at least three alternatives a social choice rule satisfies unrestricted domain, strong Pareto condition, IIA, and has transitive explanations. If for this rule a set T is locally decisive for one alternative against another, then T is golbally decisive between any two alternatives.

Proof: Suppose that x D_T y; we wish to show that z D_T* w, were z and w are any two alternatives in **O**. The proof works in two parts--first show that broad contagion holds over any triple of alternatives and then use this to show that it holds over all of **O**.

5. **Proof of Arrow's Theorem**: Assume that all five conditions hold. Now we will show that this implies a contradiction.

a. By the strong Pareto condition, there exist decisive sets.

(1) Let T be a decisive set of the smallest size (if there are more than one, just pick one arbitrarily).

(2) By the no dictator condition, T must have at least two members.

(3) From T choose one member, who will be called k.

(4) T - k is still non-empty.

(5) Consider the following profile (which should be familiar from the Condorcet paradox), **R**:

(a) k:	xyz
(b) T-k:	yzx
(c) N-T:	zxy

b. The contradiction we are after takes the form of showing that k must be a dictator.

(1) By the broad contagion result, it is sufficient to show that k is locally decisive for x against z.

(2) By IIA is would suffice to show that, at **R**, $C_{\mathbf{R}}(\{x,z\}) = \{x\}$ for then x alone would be chosen from {x,z} at any profile that, like **R**,

has k strictly preferring x to z and everyone else opposed.

c. Since T is decisive, $C_{\mathbf{R}}(\{y,z\}) = \{y\}$.

(1) If $C_{\mathbf{R}}(\{x,y\}) = \{y\}$, then IIA would tell us that T-k is locally decisive for y against z.

(2) But then the broad contagion result says that T-k is decisive, contrary to our choice of T as a minimal decisive set.

(3) Hence $C_{\mathbf{R}}(\{x,y\}) \neq \{y\}$, i.e. $x \in C_{\mathbf{R}}(\{x,y\})$.

d. Now recall that $C_{\mathbf{R}}$ is explainable by a reflexive, complete, transitive Ω.

(1) From $x \in C_{\mathbf{R}}(\{x,y\})$, we get $x \,\Omega\, y$;

(2) From $y \in C_{\mathbf{R}}(\{y,z\})$ we get $y \,\Omega\, z$; and

(3) From transitivity we get $x \,\Omega\, z$, thus $x \in C_{R'}(\{x,z\})$.

e. Now we only need to show that $z \notin C_{\mathbf{R}}(\{x,z\})$.

(1) Suppose that it is. Then $z \,\Omega\, x$ which, with $x \,\Omega\, y$, tells us that $z \,\Omega\, y$, which we know is false from $C_{\mathbf{R}}(\{y,z\}) = \{y\}$.

(2) Thus, $C_{\mathbf{R}}(\{x,z\}) = \{x\}$, which implies that k is a dictator.

f. This is a contradiction, which implies that no social choice function can satisfy all five conditions ■