Economics 101A Notes on Arrow's Impossibility Theorem

Steve Goldman

Abstract. This note reproduces Sen's proof of Arrow's Theorem from *Collective Choice and Social Welfare.*

We shall consider a Social Welfare Function f: $R^1 \times R^2 \times ...R^N \to R$, where $R^i = R$ denotes the space of complete and transitive preference orderings over a set of mutually exclusive alternatives $\{x, y, z, ...\}$. The assumption that R includes all possible orderings is called **Universal Domain**.

Definitions:

If, for some group of individuals V, $\forall i \in V$, xP^iy and $\forall j \notin V$, yP^jx imply xPy, then V is said to be *almost decisive* for alternative x against alternative y (written $D_V(x, y)$).

If, for some group of individuals V, $\forall i \in V$, xP^iy implies xPy, then V is said to be *decisive* for alternative x against alternative y (written $\overline{D}_V(x, y)$).

A *dictator* is an individual J, such that $\forall R^1, R^2, ...R^N$ and $\forall x, y$

$$xP^Jy \Rightarrow xPy$$

This is only slightly weaker than saying that f is a projection map. i.e. $f(R^1, R^2, ...R^N) \equiv R^J$ (note that if J is indifferent between two outcomes, the projection map would imply social indifference whereas Sen's would not).

Assumptions:

The set of possible members for f is restricted by the following additional assumptions:

Pareto Assumption: If $\forall i \in V, xR^iy$ then xRy (or equivalently $xf(R^1, R^2, ...R^N)y$. That is, if everyone prefers x to y then x is socially preferred to y.

Independence Assumption: Consider two alternative sets of individual orderings $R^1, R^2, ...R^N$ and $\overline{R^1}, \overline{R^2}, ...\overline{R^N}$:

If for a pair of alternatives x and y, $xR^iy \longleftrightarrow x\overline{R}^iy$ and$yR^ix \longleftrightarrow y\overline{R}^ix$ then $xRy \longleftrightarrow x\overline{R}y$. That is, all that is relevant in the social ordering of x and y are the individual orderings of x and y.

Lemma 1: If there exist x and y and an individual J such that $D_J(x, y)$ then $\forall w, z$, $\overline{D}_J(w, z)$. That is, J is a dictator.

Proof: Suppose $D_J(x, y)$ and the following preferences (since f is defined for all possible R's):

xP^Jy and yP^Jz

$\forall i \neq J, yP^ix$ and yP^iz

Now $D_J(x,y)$ and xP^Jy and $\forall i \neq J, yP^ix$ imply xPy by the definition of D.

Additionally yP^Jz and $\forall i \neq J, yP^iz$ imply yPz by the Pareto Assumption.

xPy and yPz imply xPz by transitivity of f or R.

Now what is known about individual preferences regarding x and z?

xP^Jz by assumption, but nothing is known about the preferences between x and z for anyone else. By the Independence Assumption, the relative positioning of y is irrelevant and so whenever xP^Jz the outcome must be the same. Therefore, $\overline{D}_J(x,z)$ and $D_J(x,z)$. That is, whenever J prefers x to z then x is socially preferred to z.

Suppose zP^Jx and xP^Jy while $\forall i \neq J, zP^ix$ and yP^ix.

Reasoning similarly, zPx (by Pareto) and xPy (since J is almost decisive), so zPy (by transitivity).

Therefore $\overline{D}_J(z,y)$ and hence $D_J(z,y)$.

In this manner we have shown that if J is almost decisive for x against y then J is almost decisive for x against anything and for anything against y. Sequentially applying these arguments, J is decisive for anything against anything else.

Lemma 2: There must exist an almost decisive individual.

Proof: Let Vdenote the *smallest* almost decisive group, say for x against y. V exists since the entire group is trivially almost decisive. Divide V into a single individual J and the remainder $\widehat{V}$ and the remaining population (perhaps null) as W.

Suppose the following preferences:

xP^Jy and yP^Jz

$\forall i \in \widehat{V}, zP^ix$ and xP^iy

$\forall k \in W, yP^kz$ and zP^kx.

Now, xPy since everyone in V prefers x to y and everyone in W prefers y to x and $D_V(x,y)$.

If zPy then since only members of $\widehat{V}$ have these preferences and everyone else has the opposite, $\widehat{V}$ would be a smaller almost decisive group than V - a contradiction. So, since R must be complete, yRz. By transitivity this along with xPy (from $D_V(x,y)$ and the preferences xP^Jy and $\forall i \in \widehat{V}, xP^iy$) implies xPz. But only J prefers x to z while everyone else prefers the opposite. So J is almost decisive and$\widehat{V}$ must be null. Therefore the smallest almost decisive set has but one member.

Theorem: The assumptions of Universal Domain, Pareto and Independence of Irrelevant Alternatives are consistent only with Dictatorship.