Proof of the Arrow Impossibility Theorem

Notation and Definitions

We denote the set of social alternatives by S and assume that it contains at least 3 elements. We denote the finite set of individuals by N and assume that $\#N = n \geq 2$. Each individual $i \in N$ will be assumed to have a binary weak preference relation R_i over S. The asymmetric parts of binary relations R_i, R_i', R, R' etc., will be denoted by P_i, P_i', P, P' etc., respectively; and the symmetric parts by I_i, I_i', I, I' etc., respectively.

We define a binary relation R over a set S to be (i) reflexive iff $(\forall x \in S)(xRx)$, (ii) connected iff $(\forall x, y \in S)[x \neq y \rightarrow xRy \vee yRx]$, (iii) transitive iff $(\forall x, y, z \in S)[xRy \wedge yRz \rightarrow xRz]$, (iv) an ordering iff R is reflexive, connected and transitive. We denote by T the set of all orderings over S.

A social welfare function (SWF) f is a function from $D \subseteq T^n$ to T; $f : D \mapsto T$. SWF f satisfies the condition of unrestricted domain (U) iff $D = T^n$. In other words, SWF f satisfies Condition U iff the domain of f consists of all logically possible n-tuples $(R_1, \ldots, R_n)$ of orderings. The social orderings corresponding to $(R_1, \ldots, R_n)$, $(R_1', ..., R_n')$ etc., will be denoted by R, R' etc., respectively.

An SWF satisfies (i) the weak Pareto-criterion (P) iff $(\forall (R_1, \ldots, R_n) \in D)(\forall x, y \in S)[(\forall i \in N)(xP_iy) \rightarrow xPy]$, (ii) binariness or independence of irrelevant alternatives (I) iff $(\forall (R_1, \ldots, R_n), (R_1', ..., R_n') \in D)(\forall x, y \in S)[(\forall i \in N)[(xR_iy \leftrightarrow xR_i'y) \wedge (yR_ix \leftrightarrow yR_i'x)] \rightarrow [(xRy \leftrightarrow xR'y) \wedge (yRx \leftrightarrow yR'x)]]$. $j \in N$ is called a dictator iff $(\forall (R_1, \ldots, R_n) \in D)(\forall x, y \in S)[xP_jy \rightarrow xPy]$. An SWF is called dictatorial iff $(\exists j \in N)(\forall (R_1, \ldots, R_n) \in D)(\forall x, y \in S)[xP_jy \rightarrow xPy]$. An SWF satisfies the condition of non-dictatorship (D) iff it is not dictatorial.

Let $V \subseteq N$. Let $x, y \in S, x \neq y$. We define the set of individuals V to be (i) almost decisive for (x, y) $[D(x, y)]$ iff $(\forall (R_1, \ldots, R_n) \in D)[(\forall i \in V)(xP_iy) \wedge (\forall i \in N - V)(yP_ix) \rightarrow xPy]$, (ii) decisive for (x, y) $[\overline{D}(x, y)]$ iff $(\forall (R_1, \ldots, R_n) \in D)[(\forall i \in V)(xP_iy) \rightarrow xPy]$, (iii) decisive iff it is decisive for every $(a, b) \in S \times S, a \neq b$.

$V \subseteq N$ is defined to be a minimal decisive set iff it is a decisive set and no proper subset of it is a decisive set.

Lemma : Let the social welfare function $f : T^n \mapsto T$ satisfy independence of irrelevant alternatives and the weak Pareto criterion. Then, whenever a group of individuals $V \subseteq N$ is almost decisive for some ordered pair of distinct alternatives, it is decisive for every ordered pair of distinct alternatives.

Proof : Let V be almost decisive for (x, y), $x \neq y$, $x, y \in S$. Let z be an alternative distinct from x and y, and consider the following configuration of individual preferences:
$(\forall i \in V)[xP_iy \wedge yP_iz]$
$(\forall i \in N - V)[yP_ix \wedge yP_iz]$.
In view of the almost decisiveness of V for (x, y) and the fact that $[(\forall i \in V)(xP_iy) \wedge (\forall i \in N - V)(yP_ix)]$, we obtain xPy. From $(\forall i \in N)(yP_iz)$ we conclude yPz, by condition P. From

xPy and yPz we conclude xPz, by transitivity of R. As $(\forall i \in V)(xP_iz)$, and the preferences of individuals in $N - V$ have not been specified over $\{x, z\}$, it follows, in view of condition I, that V is decisive for (x, z). Similarly, by considering the configuration
$(\forall i \in V)(zP_ix \wedge xP_iy)$
$(\forall i \in N - V)(zP_ix \wedge yP_ix)$
we can show $[D(x, y) \rightarrow \overline{D}(z, y)]$. By appropriate interchanges of alternatives it follows that $D(x, y) \rightarrow \overline{D}(a, b)$, for all $(a, b) \in \{x, y, z\} \times \{x, y, z\}$, where $a \neq b$. To prove the assertion for any $(a, b) \in S \times S, a \neq b$, first we note that if $[(a = x \vee a = y) \vee (b = x \vee b = y)]$, the desired conclusion $\overline{D}(a, b)$ can be obtained by considering a triple which includes all of x, y, a and b. If both a and b are different from x and y, then one first considers the triple $\{x, y, a\}$ and deduces $\overline{D}(x, a)$ and hence $D(x, a)$, and then considers the triple $\{x, a, b\}$ and obtains $\overline{D}(a, b)$.

Theorem: There does not exist any SWF satisfying conditions U, P, I and D.

Proof: By Condition P, N is a decisive set. Let $V \subseteq N$ be a minimal decisive set. By condition P, V is nonempty. By Condition D, $\#V \geq 2$. Let (V_1, V_2) be a partition of V [i.e., $V_1 \neq \emptyset, V_2 \neq \emptyset, V_1 \cap V_2 = \emptyset, V_1 \cup V_2 = V$]. Consider the following configuration of individual preferences:
$(\forall i \in V_1)[xP_iyP_iz]$
$(\forall i \in V_2)[yP_izP_ix]$
$(\forall i \in N - V)[zP_ixP_iy]$.
From $(\forall i \in V)(yP_iz)$, we obtain yPz.
$yPx \vee xRy$, as R is connected
$yPx \rightarrow V_2$ is almost decisive for (y, x)
$\rightarrow V_2$ is a decisive set
This contradicts the minimality of V.
$xRy \rightarrow xPz$, by transitivity of R in view of yPz
$\rightarrow V_1$ is almost decisive for (x, z)
$\rightarrow V_1$ is a decisive set
This contradicts the minimality of V.
The theorem is established as both yPx and xRy lead to contradiction.