A Proof of Gibbard Theorem

1 Notation, Definitions and Assumptions

We denote the nonempty finite set of individuals by $N = \{1, \ldots, n\}$, $n \geq 2$; and the nonempty finite set of outcomes by S, $\#S \geq 3$. Let $S \subseteq Z$. Let T denote the set of reflexive, connected and transitive binary relations (orderings) on Z, and L the set of reflexive, connected, anti-symmetric and transitive binary relations (strict orderings) on Z. We denote by C the set of all binary relations on S, and by W the set of all reflexive, connected and transitive binary relations on S. For any binary relation R we denote its asymmetric part by $P(R)$ and symmetric part by $I(R)$. For $R \in T$ and $A \subseteq Z$, we denote the restriction of R to A by $R|A$. A social choice function or voting scheme v is a surjective (onto) function from T^n to S; $v : T^n \mapsto S$. v is nonmanipulable iff $(\forall(R_1, \ldots, R_n) \in T^n)(\forall i \in N)(\forall R'_i \in T)[v(R_1, \ldots, R_n)\ R_i\ v(R_1, \ldots, R'_i, \ldots, R_n)]$. For $R \in T$ and $A \subseteq Z$, we define the set of best elements in A according to R, to be denoted by $C(A, R)$, by $\{x \in A \mid (\forall y \in A)(xRy)\}$. v is dictatorial iff $(\exists j \in N)(\forall(R_1, \ldots, R_n) \in T^n)[v(R_1, \ldots, R_n) \in C(S, R_j)]$.

Let $Q \in L$ be a fixed strict ordering such that $(\forall x, y \in Z)[x \in S \wedge y \in Z - S \rightarrow xQy]$. We define, for $R \in T$ and $A \subseteq S$, $R * A \in L$ by: (i) $R * A|Z - A = Q|Z - A$ (ii) $(\forall x, y \in Z)[x \in A \wedge y \in Z - A \rightarrow xR * Ay]$ (iii) $(\forall \text{ distinct } x, y \in A)[xR * Ay \text{ iff } xPy \vee (xIy \wedge xQy)]$.

Corresponding to $(R_1, \ldots, R_n) \in T^n$, we define binary relation $g(R_1, \ldots, R_n)$ on S as follows: $(\forall x, y \in S)[xg(R_1, \ldots, R_n)y \text{ iff } x = y \vee y \neq v(R_1 * \{x, y\}, \ldots, R_n * \{x, y\})]$.

Remark 1 *Let* $(R_1, \ldots, R_n) \in W^n$ *and* $(R'_1, \ldots, R'_n), (R''_1, \ldots, R''_n) \in T^n$ *be such that* $(\forall i \in N)(R_i = R'_i \mid S = R''_i \mid S)$. *Then* $(\forall i \in N)(\forall x, y \in S)[R'_i * \{x, y\} = R''_i * \{x, y\}]$. *For* $R_i \in W$, *let* $T_e(R_i) \subset T$ *denote the set of all extensions of* R_i *to* Z. *Thus we have shown that* $(\forall R'_i, R''_i \in T_e(R_i))[R'_i * \{x, y\} = R''_i * \{x, y\}]$, $i \in N$. *Consequently we have, for* $(\forall(R_1, \ldots, R_n) \in W^n)(\forall(R'_1, \ldots, R'_n), (R''_1, \ldots, R''_n) \in T^n)[(R'_1, \ldots, R'_n), (R''_1, \ldots, R''_n) \in T_e(R_1) \times \ldots \times T_e(R_n) \rightarrow g(R'_1, \ldots, R'_n) = g(R''_1, \ldots, R''_n)]$.

Let $f : W^n \mapsto C$ be defined by: $(\forall(R_1, \ldots, R_n) \in W^n)[f(R_1, \ldots, R_n) = R = g(R'_1, \ldots, R'_n]$, where for each $i \in N$, R'_i is some extension of R_i.

2 Manipulability of Voting Schemes

Theorem (Gibbard): If v is a nonmanipulable social choice function then v is dictatorial.

Proof: Let $\emptyset \neq A \subset S$, $(R_1, \ldots, R_n) \in T^n$, and $(\forall i \in N)(\forall x \in A)(\forall y \in S-A)(xP_iy)$. Suppose $v(R_1, \ldots, R_n) \in S-A$. By the definition of v there exists $(R'_1, \ldots, R'_n) \in T^n$ such that $v(R'_1, \ldots, R'_n) \in A$. Rename $(R_1, \ldots, R_n)$ as $(R^0_1, \ldots, R^0_n)$. Construct $(R^t_1, \ldots, R^t_n), t = 1, \ldots, n$, as follows: $[(\forall i \in N - \{t\})(R^t_i = R^{t-1}_i) \wedge R^t_t = R'_t]$. Let k be the smallest integer such that $v(R^k_1, \ldots, R^k_n) \in A$. Thus we have $v(R'_1, \ldots, R'_{k-1}, R_k, \ldots, R_n) \in S - A$, $v(R'_1, \ldots, R'_{k-1}, R'_k, \ldots, R_n) \in A$, and $v(R'_1, \ldots, R'_{k-1}, R'_k, \ldots, R_n)\ P_k\ v(R'_1, \ldots, R'_{k-1}, R_k, \ldots, R_n)$, implying that v is manipulable.
Therefore: $[\emptyset \neq A \subset S \wedge (R_1, \ldots, R_n) \in T^n \wedge (\forall i \in N)(\forall x \in A)(\forall y \in S - A)(xP_iy)] \rightarrow v(R_1, \ldots, R_n) \in A$. (A1)

By the definition of R and (A1), we conclude:
R is reflexive, connected and anti-symmetric. (A2)

Let $(R_1, \ldots, R_n) \in L^n$ and $v(R_1, \ldots, R_n) = x$. Suppose $(\exists y \in S - \{x\})(yPx)$. Rename $(R_1, \ldots, R_n)$ as $(R^0_1, \ldots, R^0_n)$ and $(R_1 * \{x, y\}, \ldots, R_n * \{x, y\})$ as $(R'_1, \ldots, R'_n)$. As $(R_1, \ldots, R_n) \in L^n$, we have $(\forall i \in N)[R_i|\{x, y\} = R'_i|\{x, y\}]$. Construct $(R^t_1, \ldots, R^t_n), t = 1, \ldots, n$, as follows: $[(\forall i \in N - \{t\})(R^t_i = R^{t-1}_i) \wedge R^t_t = R'_t]$. Let k be the smallest integer such that $v(R^k_1, \ldots, R^k_n) = z \neq x$. If $[z \in S - \{x, y\} \vee (z = y \wedge xP_ky \wedge xP'_ky)]$ then $v(R'_1, \ldots, R'_{k-1}, R_k, \ldots, R_n)P'_kv(R'_1, \ldots, R'_{k-1}, R'_k, \ldots, R_n)$, implying that v is manipulable. On the other hand, if $(z = y \ \wedge\ yP_kx \ \wedge\ yP'_kx)$ then $v(R'_1, \ldots, R'_{k-1}, R'_k, \ldots, R_n)\ P_k\ v(R'_1, \ldots, R'_{k-1}, R_k, \ldots, R_n)$, again implying that v is manipulable. Consequently, in view of (A2), we obtain:
$(R_1, \ldots, R_n) \in L_n \wedge v(R_1, \ldots, R_n) = x \rightarrow (\forall y \in S - \{x\})(xPy)$. (A3)

Let $(R_1, \ldots, R_n), (R'_1, \ldots, R'_n) \in W^n$, $x, y \in S$ and $(\forall i \in N)[R_i|\{x, y\} = R'_i|\{x, y\}]$. For $i \in N$, let $R^1_i \in T$ be some extension of R_i to Z; and $R^2_i \in T$ some extension of R'_i to Z. Then $(\forall i \in N)[R^1_i * \{x, y\} = R^2_i * \{x, y\}]$. Consequently $R|\{x, y\} = R'|\{x, y\}$. Thus:
f satisfies the condition of independence of irrelevant alternatives. (A4)

Suppose $(R_1, \ldots, R_n) \in W^n$, $x, y \in S$ and $(\forall i \in N)(xP_iy)$. For $i \in N$, let $R^1_i \in T$ be some extension of R_i to Z. Then $v(R^1_1 * \{x, y\}, \ldots, R^1_n * \{x, y\}) = x$ by (A1). We obtain xPy by the definition of R. Therefore:
f satisfies the weak Pareto criterion (A5)

Suppose $(R_1, \ldots, R_n) \in W^n$, $x, y, z \in S$, and xRy and yRz. For $i \in N$, let $R^1_i \in T$ be

some extension of R_i to Z. If $x = y$ then xRz from yRz; if $y = z$ then xRz from xRy; and if $x = z$ then xRz from reflexivity of R. Let x, y, z be all distinct. Then from $(xRy \wedge yRz)$ we obtain $xPy \wedge yPz$ by (A2). Let $(R_1^1 * \{x, y, z\}, \ldots, R_n^1 * \{x, y, z\}) = (R_1', \ldots, R_n') \in L^n$. Note that $(\forall a, b \in x, y, z)[aPb \text{ iff } aP'b]$. By (A1), $v(R_1', \ldots, R_n') \in \{x, y, z\}$. If $v(R_1', \ldots, R_n') = z$ then by (A3) $zP'y$ and therefore zPy, contradicting yPz of the hypothesis. Therefore $z \neq v(R_1', \ldots, R_n')$. Similarly, $v(R_1', \ldots, R_n') = y$ will contradict xPy and therefore $y \neq v(R_1', \ldots, R_n')$. Consequently $x = v(R_1', \ldots, R_n')$, which by (A3) implies $xP'z$, and thus xPz. Therefore:
R is transitive. (A6)

(A2), (A6), (A4) and (A5) establish that f is an Arrow social welfare function satisfying the independence of irrelevant alternatives and the weak Pareto criterion. f satisfies the condition of unrestricted domain by definition. f is therefore dictatorial, i.e., $(\exists j \in N)(\forall (R_1, \ldots, R_n) \in W^n)(\forall x, y \in S)[xP_jy \rightarrow xPy]$. (A7)

Let j be dictator for f. Suppose $v(R_1, \ldots, R_n) = x$ and $x \notin C(S, R_j)$. Let $y \in C(S, R_j)$ and thus yP_jx. Let $R_j' \in L$ be such that $(\forall z \in Z - \{y\})(yP_j'z)$; and for each $i \neq j, R_i' \in L$ be such that $(\forall z \in Z - \{y\})(zP_i'y)$. By (A3) and the fact that j is a dictator for f, $v(R_1', \ldots, R_n') = y$. Suppose $(\exists R_1'', R_2'', \ldots, R_{j-1}'', R_{j+1}'', \ldots, R_n'' \in T)$ $[v(R_1'', R_2'', \ldots, R_{j-1}'', R_j', R_{j+1}'', \ldots, R_n'') = z \neq y]$. Rename $(R_1', \ldots, R_n')$ as $(R_1^0, \ldots, R_n^0)$. Construct $(R_1^t, \ldots, R_n^t)$, $t = 1, \ldots, j-1, j+1, \ldots, n$, as follows: $(\forall t \in \{1, \ldots, j-1, j+2, ..., n\})[R_j^t = R_j' \wedge R_t^t = R_t'' \wedge (\forall i \in N - \{t, j\})(R_i^t = R_i^{t-1}) \wedge$ $[R_j^{j+1} = R_j' \wedge R_{j+1}^{j+1} = R_{j+1}'' \wedge$ $(\forall i \in N - \{j, j+1\})[R_i^{j+1} = R_i^{j-1}]$. Let k be the smallest integer such that $v(R_1^k, \ldots, R_n^k) \neq y$. Then $v(R_1^k, \ldots, R_n^k)\ P_k^{k-1}\ v(R_1^{k-1}, \ldots, R_n^{k-1})$ or $v(R_1^k, \ldots, R_n^k) P_k^{k-2}\ v(R_1^{k-2}, \ldots, R_n^{k-2})$ implying that v is manipulable. Therefore, $(\forall R_1'', R_2'', \ldots, R_{j-1}'', R_{j+1}'', \ldots, R_n'' \in T)$ $[v(R_1'', R_2'', \ldots, R_{j-1}'', R_j', R_{j+1}'', \ldots, R_n'') = y]$. Consequently, $v(R_1, \ldots, R_{j-1}, R_j', R_{j+1}, \ldots, R_n)$ $P_j\ v(R_1, \ldots, R_n)$, implying that v is manipulable. Therefore, $[v(R_1, \ldots, , R_n) = x \rightarrow x \in C(S, R_j)]$. This establishes that j is dictator for v.

Satish K. Jain
Centre for Economic Studies and Planning
School of Social Sciences
Jawaharlal Nehru University
New Delhi

satish.k.jain@gmail.com

September 20, 2005