Comparing impossibility theorems

Randy Calvert, for Pol Sci 507 Spr 2017

All references to "A-S & B" are to Austen-Smith and Banks (1999).

Basic notation

- X set of alternatives
- $\mathcal{X}$ set of all nonempty subsets of alternatives
- N set of individuals, with $|N| = n < \infty$
- $\mathcal{B}$ set of all complete, reflexive binary relations on X
- $\mathcal{R}$ set of all transitive, complete, reflexive binary relations (i.e. weak orders) on X; typical element R
- P, I the asymmetric and symmetric part, resp., of R; similarly for P_i, I_i, R_i
- $\mathcal{R}^n$ set of all profiles of weak orders; typical element $\rho = (R_1, \ldots, R_n)$
- $F : \mathcal{D} \to \mathcal{B}$, where $\mathcal{D} \subseteq \mathcal{R}^N$: a **preference aggregation rule**, or **social welfare function**
  - Implicitly, for every $\rho \in \mathcal{D}$, $F(\rho)$ is complete and reflexive. Substantively, this means that rather than ever saying x and y are incomparable, we say that socially xIy.
- $\phi : \mathcal{X} \times \mathcal{D} \to \mathcal{X}$ with $\phi(S, \rho) \subseteq S$ is a **social decision function**
  - If a preference aggregation function $F(\rho)$ is acyclic (i.e. has no strict-preference cycles of any length) for every $\rho \in \mathcal{D}$ then it implies a social decision function $\phi(S, \rho) \stackrel{\text{def}}{=} \max(S, F(\rho))$, the $F(\rho)$-maximal elements of S.
  - Conversely, if a social decision function ϕ satisfies certain regularity conditions (often called condition α and condition γ—see Austen-Smith and Banks 1999, pp. 8-11), then its *base relation* is the unique preference aggregation function that generates ϕ. The base relation is defined by $xRy \Leftrightarrow x \in \phi(\{x, y\}, \rho)$.
- $f : \mathcal{D} \to \mathcal{X}$ is a **social choice function**.
  - Given a social decision function ϕ, $f(\rho) = \phi(X, \rho)$ is a social choice function.
  - Conversely under suitable conditions–for example if f satisfies universal domain and social-choice-function versions of IIA, and Pareto—a social choice function f can be used to generate a full social decision function ϕ by the following operation: for any ρ, form $\rho'(x, y)$ by moving x and y to the top of each individual's preference ranking, maintaining each individual's relative ranking of x and y. Then define R by $xRy \Leftrightarrow x \in f(\rho'(x, y))$; and define $\phi(S, \rho) \stackrel{\text{def}}{=} M(S, R)$.

Conditions on F or on ϕ

- universal domain: $\mathcal{D} = \mathcal{R}^n$
- transitive social preference: for every $\rho \in \mathcal{D}$, $F(\rho)$ is transitive
- Pareto (Blau's UPP): For any ρ, if xP_iy for every $i \in N$ then xPy socially (that is, $xf(\rho)y$ and not $yf(\rho)x$). (A-S&B call this condition "weak Pareto.")
- weak Pareto (Blau's UPR): For any ρ, if xP_iy for every $i \in N$ then xRy socially (that is, $xf(\rho)y$)
- independence of irrelevant alternatives (IIA): For every $\rho, \rho' \in \mathcal{D}$,

$$\left[\forall i \in N, \forall x, y \in X \;\; xR_iy \Leftrightarrow xR'_iy\right] \Rightarrow F(\rho) = F(\rho').$$

- nondictatorship: for every $i \in N$, there exist ρ, x, y such that xP_iy but not xPy.
- non-null (Blau): there exists $\rho \in \mathcal{D}$ such that $F(\rho) \neq I$, that is,

$$\exists \rho \in \mathcal{D}, x, y \in X \text{ such that not } xF(\rho)y$$

- neutrality (Blau): Assume universal domain. For every $\rho \in \mathcal{R}^n$, for every permutation $\sigma : X \to X$, for each $i \in N$ define R'_i by

$$xR'_iy \Leftrightarrow \sigma(x)R_i\sigma(y)$$

  Let $R = F(R_1, \ldots, R_n)$ and $R' = F(R'_1, \ldots, R'_n)$. Then $xRy \Leftrightarrow \sigma(x)R'_i\sigma(y)$.

  - Note: a version of neutrality without the assumption of universal domain could be stated by requiring instead that for every $\rho \in \mathcal{D}$ and every σ, $(R_1, \ldots, R_n) \in \mathcal{D} \Rightarrow (R'_1, \ldots, R'_n) \in \mathcal{D}$.

- nonmanipulability (Gibbard; elsewhere, "incentive compatibility"): This condition is stated in terms of a single-valued social choice function $f : \mathcal{D} \to X$. Assume universal domain. For every $\rho = (R_1, \ldots, R_n) \in \mathcal{R}^n$, for every $i \in N$, for every $x, y \in X$, and for every $R'_i \in \mathcal{R}$,

$$f(\rho) \; R_i \; f((R'_i, R_{-i})).$$

  - Note: a version of nonmanipulability without the assumption of universal domain could be stated by applying the condition only to $\rho \in \mathcal{D}$ and to R'_i such that $(R'_i, R_{-i}) \in \mathcal{D}$.

- liberalism (Amartya Sen, 1970): For each $i \in N$, there is at least one pair of alternatives, say (x_i, y_i), such that for every ρ, the social preference $R = F(\rho)$ satisfies $x_iP_iy_i \Rightarrow x_iPy_i$ and $y_iP_ix_i \Rightarrow y_iPx_i$. (By implication, liberalism also requires X to be large enough to accommodate the required x_i and y_i.)

  - Sen's theorem actually requries only the weaker condition of "minimal liberalism": there are at least two distinct individuals $i, j \in N$ for which the above condition holds.

Arrow's theorem

As stated by Blau (1972, p. 63): If $|X| \geq 3$, then the following conditions cannot all hold:

- transitivity
- universal domain
- Pareto
- IIA
- nondictatorship

In his Theorem 4, Blau proves as well that Pareto can be replaced by the combination of weak Pareto and non-null.

Gibbard's theorem

For a setting in which a social choice rule ϕ is always single-valued (that is, $|\phi(S, \rho)| = 1$ for every ρ), Gibbard (1973) proves the result by showing that nonmanipulability implies that the conditions other than nondictatorship are all satisfied; hence a SCF is either dictatorial or manipulable. The theorem can thus be restated as follows: if $|X| \geq 3$ (and finite, which is unstated but needed by Gibbard) then the following conditions cannot all hold:

- transitivity
- universal domain (actually, only the weaker version below is needed)
- Pareto
- nondictatorship
- ϕ is nonmanipulable on $\mathcal{D}$

Evidently, nonmanipulability does the work of IIA. (Gibbard actually proves his result in terms of the social choice function f, but it is easily extended to a SDF under the other conditions given.)

Benoit (2000) and Arunava Sen (2001) both offer proofs of a Gibbard-type theorem using the general approach of Geanakoplos (they cite 1996; published version 2005). Their approach uses slightly weaker versions of all the Gibbard conditions, mainly reflecting their recognition that universal domain could be weakened to require only that all profiles of strict orders be in the domain. Suppose $|X| \geq 3$ and finite. Then the following conditions cannot all hold:

- quasi-transitivity (that is, transitivity of the social P-relation; a weakened form of transitivity)
- a weaker version of universal domain: every profile of asymmetric preferences is in $\mathcal{D}$
- a weaker version of Pareto: if for every $i \in N$, xP_iy, then $\phi(X, \rho) = \{x\}$

- a weaker version of nondictatorship: a dictator is an individual j such that $xP_jy \; \forall y \Rightarrow \phi(X, \rho) = \{x\}$
- ϕ is nonmanipulable on $\mathcal{D}$

(Hence, as an obvious corollary, one of the strong versions of those conditions must also be violated.)

Sen's theorem on impossibility of a Paretian liberal

Sen (1970): No preference aggregation function or social decision function can satisfy all of the following:

- transitivity
- universal domain
- Pareto
- minimal liberalism

Evidently minimal liberalism does the work of both IIA and nondictatorship.

Impossibility theorems that weaken transitivity by strengthening nondictatorship

Regularity conditions on R weaker than transitivity:

- quasi-transitivity: A binary relation R on X is *quasi-transitive* iff xPy and $yPz \Leftrightarrow xPz$. This is sometimes informally called P-transitivity or, in Sen (1970), "PP."
- A SDF is quasi-transitive iff its base relation is quasi-transitive.
- acyclicity: A binary relation R on X is *acyclic* iff there is no cycle under P, that is, no sequence of alternatives $x_1, x_2, \ldots, x_k$ such that x_kPx_1 and, for every $i \in \{1, 2, \ldots, k-1\}$, x_iPx_{i+1}.
- A SDF is acyclic iff its base relation is acyclic. Since an SDF $\phi(S, \rho)$ is assumed nonempty for every $S \in \mathcal{X}$ and every $\rho \in \mathcal{D}$, its base relation—and hence any R that rationalizes it—is acyclic (follows from A-S & B, Theorem 1.2).

Decisive coalitions for a SWF F and for an SDF ϕ:

- A coalition, or set of individuals, $C \subseteq N$ is *decisive under* F if $\left[\forall \rho \in \mathcal{D} \text{ and } \forall i \in C \; xP_iy\right] \Rightarrow xPy$ where $P = F(\rho)$.
- A coalition C is *decisive under* ϕ if $\left[\forall \rho \in \mathcal{D} \text{ and } \forall i \in C \; xP_iy\right] \Rightarrow \phi(\{x, y\}, \rho) = \{x\}$.
- A set $C \subseteq N$ is an *oligarchy* iff for every decisive set D, $C \subseteq D$.

  - This means that every member i of C “has veto power” in the sense that $\forall x, y \in X\ \big[xP_i y \Rightarrow \neg(yPx)\big]$. In terms of the SDF, $xP_i y \Rightarrow \big[x \in S \Rightarrow y \notin \phi(S,\rho)\big]$.
  - In this sense, an oligarchy is the set of C individuals having veto power, provided C itself is decisive.
  - A dictator is a singleton oligarchy.

- The *collegium* is the intersection of all decisive sets under F or under ϕ. If its collegium is nonempty, F or ϕ js said to be collegial.
  - An oligarchic rule is collegial.
  - A member of a collegium need not have veto power (Example 2.6 in A-S & B)

Theorem (Gibbard 1969 unpublished; Guha 1972, Mas-Colell and Sonnenschein 1972)). No preference aggregation function can satisfy all of the following

- quasi-transitivity
- universal domain
- Pareto
- IIA
- there is no oligarchy

Theorem (Brown 1975). No preference aggregation function can satisfy all of the following

- acyclicity
- universal domain
- Pareto
- there is no collegium

This theorem does not require IIA. However, as A-S & B point out, notice that the existence of decisive coalitions (at least one, due to Pareto) implies at least a limited form of independence.

References

Kenneth J. Arrow, *Social Choice and Individual Values* (Wiley, 1951); Second edition (Yale Univ. Press 1963).

David Austen-Smith and Jeffrey S. Banks, *Positive Political theory I: Collective Preference.* Univ. of Michigan Press, 1999.

Jean-Pierre Benoit, "The Gibbard-Satterthwaite theorem: a simple proof." *Economics Letters* Vol. 69 (2000), pp 319-22.

Julian H. Blau, "A Direct Proof of Arrow's Theorem." *Econometrica*, Vol. 40, No. 1 (Jan., 1972), pp. 61-67.

Donald J. Brown, "Aggregation of Preferences." *Quarterly Journal of Economics* Vol. 89 (1975), pp. 456-69.

John Geanakoplos, "Three Brief Proofs of Arrrow's Impossibility Theorem." *Economic Theory*, Vol. 26 (2005), pp. 211-215.

Allan Gibbbard. "Social Choice and the Arrow Conditions." Discussion Paper, Dept of Philosophy, Univ. of Michigan (1969)

Allan Gibbard, "Manipulation of voting schemes." *Econometrica*, Vol. 41 (1973), pp. 587-601.

A.S. Guha, "Neutrality, Monotonicity, and the Right of Veto." *Econometrica* Vol 41 (1972), pp. 821-26.

Amartya Sen, "The Impossibility of a Paretian Liberal." *Journal of Political Economy* Vol. 78, No. 1 (Jan.-Feb., 1970), pp. 152-157.

Arunava Sen, "Another direct proof of the Gibbard-Satterthwaite Theorem." *Economics Letters* Vol. 70 (2001), pp. 381-85.

Andreu Mas-Colell and Hugo F. Sonnenschein, "General Possibility Theorems for Group Decisions." *Review of Economic Studies* Vol 39 (1972), pp. 185-92.