The Condorcet Paradox
and the Impossibility of Nice Social Choice

Randy Calvert, for Pol Sci 507 Spr 2017

Majority rule cycles

Suppose each of three individuals $\{1, 2, 3\}$ ranks three alternatives $\{a, b, c\}$ from best to worst as follows:

1	2	3
a	b	c
b	c	a
c	a	b

Define P to be a (strict) group preference ranking over X determined, on each pair of alternatives, by the preferences of a majority of individuals: the group strictly prefers x to y if and only if more individuals strictly prefer x to y than strictly prefer y to x. (We would say that the group is indifferent between x and y if neither xPy nor yPx.) The individual preferences shown above then create a "majority rule cycle" in the group preference relation. In other words, simple majority preference fails the condition of transitivity (indeed, fails the even weaker condition of acyclicity), even though all individual orderings are transitive. This example is sometimes called the "Condorcet paradox" after one of its several discoverers, the 18th-century Marquis de Condorcet.

This paradox of the majority cycle can be generalized to any set N of three or more voters, partitioned into three subgroups N_1, N_2, N_3, no one of which is a majority of N, but every pair of them comprising a majority. Suppose that for alternatives a, b, c, individuals in N_1 have the preferences of individual 1 above, those in N_2 have the preferences of individual 2, and those in N_3 have the preferences of individual 3. Then the majority preference relation exhibits a majority-rule cycle. The paradox can also be constructed with any larger number of alternatives.

Preference aggregation functions

Let X be a set of alternatives and $N = \{1, \ldots, n\}$ a set of individuals. Suppose each individual i has a transitive, reflexive, complete preference ordering R_i over X, representing "preference or indifference," that is, weak preference. Call $\rho = (R_1, \ldots, R_n)$ a group preference *profile*. Let F be a function that, for any such profile of preferences, specifies a group preference relation $R = F(\rho)$ on X; we assume only that R is a complete, reflexive binary relation.

Given any profile ρ, say that individual i strictly prefers x to y, and write xP_iy, if xR_iy but not yR_ix. An individual may also be indifferent between x and y—xI_iy—if both xR_iy

and yR_ix.

Say that F *respects majority preferences* if whenever a larger number of individuals strictly prefer x to y than strictly prefer y to x, the social preference puts xPy. Technically,

$$|\{i \in N : xP_iy\}| > |\{i \in N : yP_ix\}| \Rightarrow xF(\rho)y \text{ and not } yF(\rho)x.$$

Technically, denote by $\mathcal{B}$ the set of all complete, reflexive binary relations on X, and by $\mathcal{R}$ the set of all transitive, complete, reflexive binary relations on X. The Cartesian product $\mathcal{R}^n$ then represents the set of all preference profiles. A function $F : \mathcal{R}^n \to \mathcal{B}$, then, is called a "preference aggregation rule" (Austen-Smith and Banks 1999) or a "social welfare function" (Arrow 1951).

When we formulate a preference aggregation rule in this way, $F : \mathcal{R}^n \to \mathcal{B}$, we are assuming in advance that it always gives a social preference between every pair of alternatives (since the relations in $\mathcal{B}$ are complete). Rather than ever saying that x and y are not socially comparable, we will always say instead that society is indifferent between them. This turns out arguably to be not a strong assumption, but more a matter of nomenclature. The further assumption that every $R = F(\rho)$ is reflexive is truly a matter of consistent nomenclature: it says we will use the complete ordering R to indicate indifference as well as strict preference. We could do it otherwise: if instead we treated $F(\rho)$ as being a strict (that is, asymmetric) social ordering, $P = F(\rho)$, then since the ordering is complete, we would call society indifferent between alternatives, xIy, whenever neither xPy nor yPx. But then we could not say that P is complete Once we have assumed completeness, reflexivity just amounts to a naming convention.

An impossibility theorem

We can now state our first theorem concerning the impossibility of nice social choice.

Theorem 1 *If X contains at least three alternatives and N contains at least three individuals, there is no preference aggregation rule F that satisfies both the following conditions:*

- *For every $\rho \in \mathcal{R}^n$, the social preference relation $R = F(\rho)$ is* ***transitive****.*
- *For every $\rho \in \mathcal{R}^n$, the social preference relation $R = F(\rho)$* ***respects majority preferences****.*

A proof of this theorem results simply from applying the Condorcet paradox. If there are $a, b, c \in X$ on which the profile of individual preferences matches that given in the paradox, then there is not way to assign a group preference over those three alternatives without violating either transitivity or majority rule. For majority rule requires aPb and bPc under the strict group preference P; transitivity then requires cPa, which fails to respect the majority preference between a and c.

May's Theorem: an axiomatic characterization of simple majority rule

In a classic paper, Kenneth O. May (1952) defines the social aggregation rule of "simple majority rule" (SMR; this is what Austen-Smith and Banks, 1999, call the "plurality rule"). It is the preference aggregation rule defined as follows: for every $\rho \in \mathcal{R}^n$ and for every pair of alternatives $x, y \in X$,

$$xF(\rho)y \Leftrightarrow |\{i \in N : xP_iy\}| \geq |\{i \in N : yP_ix\}|.$$

May's great accomplishment was to demonstrate that SMR is precisely characterized by a short list of axioms.

Theorem 2 (May's Theorem) *If X contains at least three alternatives and N contains at least three individuals, a preference aggregation rule F is simple majority rule if and only if it satisfies all the following conditions:*

- *F satisfies* ***independence of irrelevant alternatives*** *(IIA): the if two profiles ρ and ρ' feature the same preference between x and y for every individual, then the social preference between x and y is the same under $F(\rho)$ as under $F(\rho')$.*
- *F is* ***neutral****: it is not affected by a permutation of alternatives.*
- *F is* ***anonymous****: it is not affected by a permutation of individuals.*
- *F is* ***positively responsive****: moving x "higher" relative to y is some individuals' orderings results in x doing better in social preference relative to y.*

An impossibility theorem implied by May's Theorem

Putting together Theorem 1 and May's Theorem gives an apparently more substantial impossibility theorem.

Theorem 3 *If X contains at least three alternatives and N contains at least three individuals, there is no preference aggregation rule F that satisfies all the following conditions:*

- *For every $\rho \in \mathcal{R}^n$, the social preference relation $R = F(\rho)$ is* ***transitive****.*
- *F satisfies* ***IIA****.*
- *F is* ***neutral****.*
- *F is* ***anonymous****.*
- *F is* ***positively responsive****.*

The proof: if F satisfies the last four conditions, then by May's Theorem it is SMR. But SMR respects majority preferences, so by Theorem 1 it violates transitivity for some preference profile.