Allan Gibbard - *Manipulation of voting schemes: a general result* (1973)

Charlotte Vlek

June 23, 2009

Table of contents

Allan Gibbard

- Allan Gibbard (1942 -)
- University Professor of Philosophy at University of Michigan

"My field of specialization is ethical theory"

"My current research centers on claims that the concept of meaning is a normative concept"
(www-personal.umich.edu/~gibbard/)

Situation in 1973

Conjectured: all voting schemes are manipulable.

- Dummet & Farquharson: *Stability in voting* (1961)

  > *"it seems unlikely that there is any voting procedure in which it can never be advantageous for any voter to vote "strategically", i.e., non-sincerely." (D.& F. 1961, p.34 in: Gibbard 1973, p.588)*

- They prove a similar result but only for "majority games", not for all voting schemes

Situation at the time

- Vickrey: *Utility, strategy and social decision rules* (1960):
  - IIA & positive association imply non-manipulability
  - conjectured: non-manipulability implies IIA & PA.

Gibbard confirms Vickrey

Definitions - ordering

An **ordering** of Z is two-place relation P such that for all $x, y, z \in Z$:

- $\neg(xPy \wedge yPx)$ (totality)
  (logically equivalent to $yRx \vee xRy$)
- $xPz \rightarrow (xPy \vee yPz)$ (transitivity)
  (logically equivalent to $(zRy \wedge yRx) \rightarrow zRx$)

Definitions - voting scheme

- n voters
- Z set of alternatives
- P_i orderings of Z for each voter i

A **voting scheme** is a function that assigns a member of Z to each possible **preference n-tuple** $(P_1, P_2, ..., P_n)$ for a given number n and set Z.

Definitions - manipulation

One **manipulates** the voting scheme if

> *"by misrepresenting his preferences, he secures an outcome he prefers to the "honest" outcome" (Gibbard 1973, p.587)*

Note that manipulation only has a meaning if we know the "honest" preferences too.

The main result

"Any non-dictatorial voting scheme with at least 3 possible outcomes is subject to individual manipulation" (Gibbard 1973, p. 587)

Definitions - Game form

> *"A game form is any scheme which makes an outcome depend on individual actions of some specified sort, which I shall call strategies" (Gibbard 1973, p.587)*

Formally:

- X a set of possible outcomes
- n number of players
- S_i for each player i, a set of **strategies** for i.

A **game form** is a function

$$g : S_1 \times S_2 \times \ldots \times S_n \to X$$

that takes each possible strategy n-tuple $\langle s_1, \ldots, s_n \rangle$ with $s_i \in S_i \; \forall i$ to an outcome $x \in X$.

Voting scheme vs. Game form

- Every non-chance procedure by which individual choices of contingency plans for action determine an outcome is characterized by a game form
- Voting scheme is a special case of game form
- A game form does not specify what an 'honest' strategy would be, so there is no such thing as manipulability

Voting scheme vs. Game form

- **Manipulability** is a property of a game form plus n functions σ_k $(k \leq n)$ that take each possible preference ordering to a strategy $s \in S_k$. For each individual k and preference ordering P, $\sigma_k(P)$ is the strategy for k which honestly represents P.
- Now we have

$$v(P_1, ..., P_n) = g(\sigma_1(P_1), ..., \sigma_n(P_n))$$

Definitions - dominant strategy

> *"A strategy is dominant if whatever anyone else does, it achieves his goals at least as well as would any alternative strategy" (Gibbard 1973, p.587)*

Formally:

- let $\mathbf{s} = \langle s_1, ..., s_n \rangle$ be a strategy n-tuple
- let $\mathbf{s}k/t = \langle s_1, ..., s_{k-1}, t, s_{k+1}, ..., s_n \rangle$ (replace kth strategy by t)

A strategy t is P-**dominant** for k if for every strategy n-tuple $\mathbf{s}$, $g(\mathbf{s}k/t)Rg(\mathbf{s})$.
A game form is **straightforward** if for every individual k and preference ordering P, there is a strategy P-dominant for k.

Definitions - dictatorship

- A player k is a **dictator** for a game form g if for every outcome x there is a strategy $s(x)$ for k such that $g(\mathbf{s}) = x$ whenever $s_k = s(x)$.
- A game form g is dictatorial if there is a dictator for g.

The result for game forms:
Every straightforward game form with at least three possible outcomes is dictatorial.

The result for game forms

The result for game forms:
Every straightforward game form with at least three possible outcomes is dictatorial.

Corollary:
Every voting scheme with at least three outcomes is either dictatorial or manipulable.

Proof of theorem

The result for game forms:
Every straightforward game form with at least three possible outcomes is dictatorial.

Proof:

- Let g be a straightforward game form with at least 3 outcomes
- For each i, let σ_i be such that for every P, $\sigma_i(P)$ is P-dominant for i
- Let $\sigma(\mathbf{P}) = \langle \sigma_1(P_1), ..., \sigma_n(P_n) \rangle$
- Let $v = g \circ \sigma$

Proof of theorem

- Fix some strict ordering Q. Let $Z \subseteq X$
- For each i, define $P_i * Z$ such that for all $x, y \in X$
  - If $x \in Z$ and $y \in Z$ then $x(P_i * Z)y$ iff either zP_iy or both xI_iy and xQy
  - If $x \in Z$ and $y \notin Z$ then $x(P_i * Z)y$
  - If $x \notin Z$ and $y \notin Z$ then $x(P_i * Z)y$ iff xQy
- Let $\mathbf{P} * Z = \langle P_1 * Z, \ldots, P_n * Z \rangle$
- define xPy to be

$$x \neq y \wedge x = v(\mathbf{P} * \{x, y\})$$

- Show $f(\mathbf{P}) = P$ is a social welfare function, satisfying all of Arrow's conditions except non-dicatorship
- the dictator for f is a dictator for $v = g \circ \sigma$

Implications

- Any voting scheme we use will be manipulable, unless trivial.
- Manipulability does *not* mean that in reality people are always in a position to manipulate.
  It means that it's not guaranteed that they can't.
- But reasons not to:
  - ignorance
  - integrity
  - stupidity

  But "the 'ignorance' and 'stupidity' required here are just the ordinary conditions of human existence" (Simon 2002, p. 112)

More on the subject

- This result concerns *non-chance procedures*. Mixed decision schemes can be non-manipulable. See example and Gibbard's *Manipulation of schemes that mix voting with chance*, 1977
- Correspondence Arrow's social welfare function and non-manipulable voting scheme.
  Satterthwaite:
  - Gibbard does not consider voting schemes with restricted outcomes. Can easily be fixed.
  - Gibbard does not establish uniqueness of underlying social welfare function. Easy to prove.
  - Gibbard does not prove non-negative responsiveness (NNR) for the swf. Can be done.

Discussion

- Compare Gibbard's and Satterthwaite's versions of Arrow's conditions:
  - Gibbard (p. 586): Scope; Unanimity; Pairwise Determination (equiv. to IIA); Non-dictatorship
  - Satterthwaite (p. 204): Non-dictatorship (ND); Independence of Irrelevant Alternatives (IIA); Citizen's Sovereignty (CS); Non-negative Responsiveness (NNR)
- Game forms take three steps: personal agenda $\Rightarrow$ strategy $\Rightarrow$ outcome
  Why not use this for voting schemes too: preferences $\Rightarrow$ ballot $\Rightarrow$ social choice
  (note: remember Gibbard's example with the club voting for alcoholic parties)

Literature

- Gibbard, A.; 1973. Manipulation of voting schemes: a general result. In: *Econometrica*, Vol 41, No. 4, pp.587-601.
- Dummet, M.; Farquharson, R.; 1961. Stability in Voting. In: *Econometrica*, Vol. 29, No. 1, pp. 33-43.
- Vickrey, W.; 1960. Utility, Strategy, and Social Decision Rules. In: *The Quarterly Journal of Economics*. Vol. 74, No. 4, pp. 507-535.
- Simon, R.L.; 2002. *The Blackwell guide to social and political philosophy*. Wiley-Blackwell.