Fair Division in Theory and Practice

Ron Cytron (Computer Science)
Maggie Penn (Political Science)

Lecture 7: The Gibbard-Satterthwaite Theorem

The Gibbard-Satterthwaite Theorem

Arrow's Theorem considered preference aggregation rules: rules taking a preference profile as an input and and generating a social ranking of alternatives

Today we will deal with "choice functions" that take preference profiles and generate a single outcome

Formally, a *social choice function* f maps $\mathcal{R}^n \to X$

Gibbard-Sattherwaite and Arrow deal with *preference aggregation* but in different ways

Arrow asks under what conditions can an institution take a collection of preferences as an input and produce a "rational" (transitive) social ranking of alternatives

Gibbard asks under what conditions an institution is "strategy-proof," or immune to manipulation by voters. In this case the institution takes as an input ballots that are chosen by the voters. When does a voter want to submit a ballot that insincerely represents his/her preferences?

Social Choice Functions

A *social choice function* f takes (revealed) preferences as an input and produces a single social outcome, so that $f(\rho) = x$

A social choice function is *strategy-proof* if for all individuals i, no individual can reveal an insincere preference ordering to make himself strictly better off via the social choice function

A social choice function is *dictatorial* if there exists an i such that for all revealed preference profiles ρ, x being the social choice implies that i ranked x highest (or tied for highest)

A social choice function is *onto* (or has "full range") if for every policy x, there is some collection of revealed preferences ρ so that $f(\rho) = x$

Formal definition of ***strategy-proofness*****:**

f is strategy-proof if and only if for all $\rho = (\succeq_1, ..., \succeq_i, ..., \succeq_n) \in \mathcal{R}^n$ and all $\succeq'_i \in \mathcal{R}$, it is the case that:

$$f(\rho) \succeq_i f(\succeq_1, ..., \succeq'_i, ..., \succeq_n)$$

The Gibbard-Satterthwaite Theorem

If $|X| \geq 3$ and social choice function f is onto and strategy-proof, then f is dictatorial

Like Arrow, this theorem also requires unrestricted domain – we must assume that all preference orderings are allowed

What the theorem says

Gibbard-Satterthwaite says that there is no way to generate social outcomes that can guarantee that it is never in someone's interest to misrepresent their preferences

(Strategic) insincerity on the part of voters is endemic to all (nondictatorial) political institutions

Easy examples

Plurality rule:

$$3 : Gore \succ Bush \succ Nader$$

$$3 : Bush \succ Gore \succ Nader$$

$$1 : Nader \succ Gore \succ Bush$$

Nader supporter has an incentive to submit a ballot with Gore top-ranked

What does a "ballot" mean in this context?

A ballot is simply a person's way of choosing how they want their preferences to be represented within the decision-making institution

- Could be a paper ballot, as in plurality example
- A series of votes the person makes
- Even how they choose the person that will be representing them within a larger decision-making process

A dictatorial social choice rule

	xyz	**xzy**	**yxz**	**yzx**	**zxy**	**zyx**
xyz	x	x	x	x	x	x
xzy	x	x	x	x	x	x
yxz	y	y	y	y	y	y
yzx	y	y	y	y	y	y
zxy	z	z	z	z	z	z
zyx	z	z	z	z	z	z

Change *one* cell to make it non-dictatorial

	xyz	**xzy**	**yxz**	**yzx**	**zxy**	**zyx**
xyz	x	x	x	x	x	x
xzy	x	x	x	x	x	x
yxz	y	x	y	y	y	y
yzx	y	y	y	y	y	y
zxy	z	z	z	z	z	z
zyx	z	z	z	z	z	z

Who can profit by misreporting preferences, & how?

(Hint: there are cases where either player can profit)

Change *one* cell to make it non-dictatorial

	xyz	**xzy**	**yxz**	**yzx**	**zxy**	**zyx**
xyz	x	x	x	x	x	x
xzy	x	x	x	x	x	x
yxz	y	x	y	y	y	y
yzx	y	y	y	y	y	y
zxy	z	z	z	z	z	z
zyx	z	z	z	z	z	z

Who can profit by misreporting preferences, & how?

Column player can profit by saying yellow cells are really blue

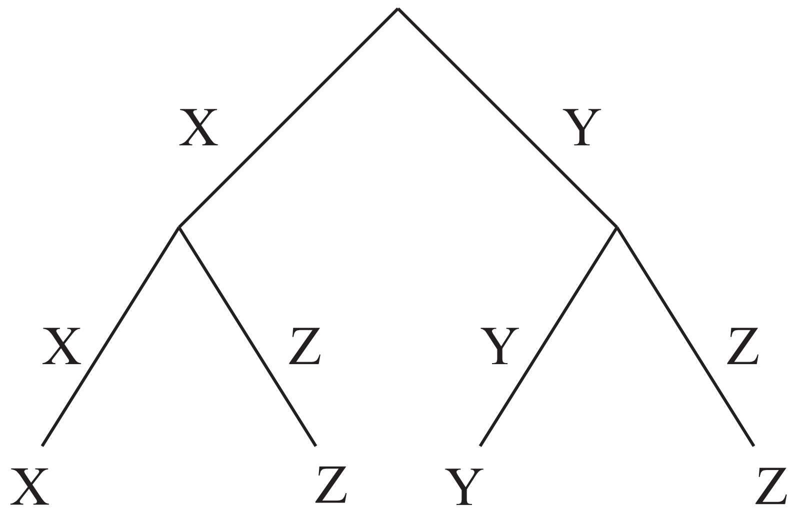
Change *one* cell to make it non-dictatorial

	xyz	**xzy**	**yxz**	**yzx**	**zxy**	**zyx**
xyz	x	x	x	x	x	x
xzy	x	x	x	x	x	x
yxz	y	x	y	y	y	y
yzx	y	y	y	y	y	y
zxy	z	z	z	z	z	z
zyx	z	z	z	z	z	z

Who can profit by misreporting preferences, & how?

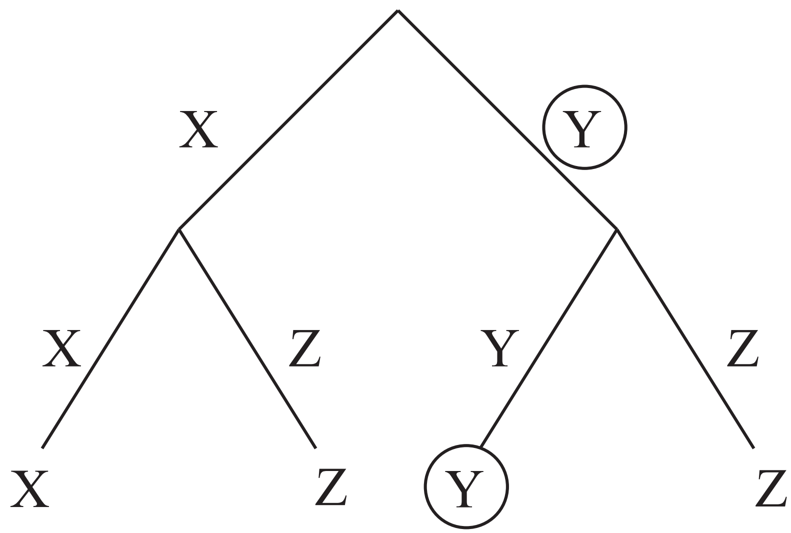
Row player can profit by saying blue cell is really yellow

An Example of Manipulation in pairwise voting



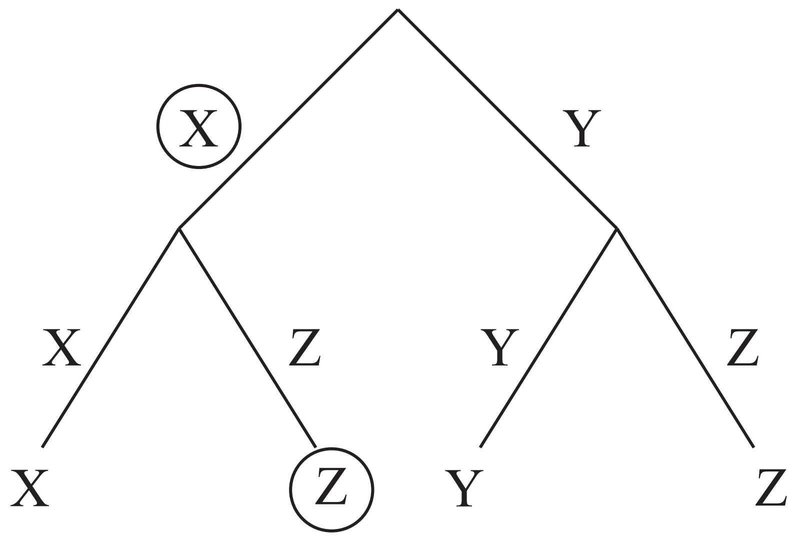
i	Sincere Preferences	Manipulation by 3
1	$X > Y > Z$	$X > Y > Z$
2	$Y > Z > X$	$Y > Z > X$
3	$Z > Y > X$	$Z > X > Y$

An Example of Manipulation in pairwise voting



i	Sincere Preferences	Manipulation by 3
1	$X > Y > Z$	$X > Y > Z$
2	$Y > Z > X$	$Y > Z > X$
3	$Z > Y > X$	$Z > X > Y$

An Example of Manipulation in pairwise voting



i	Sincere Preferences	Manipulation by 3
1	$X > Y > Z$	$X > Y > Z$
2	$Y > Z > X$	$Y > Z > X$
3	$Z > Y > X$	$Z > X > Y$

Example of manipulation in choosing a representative

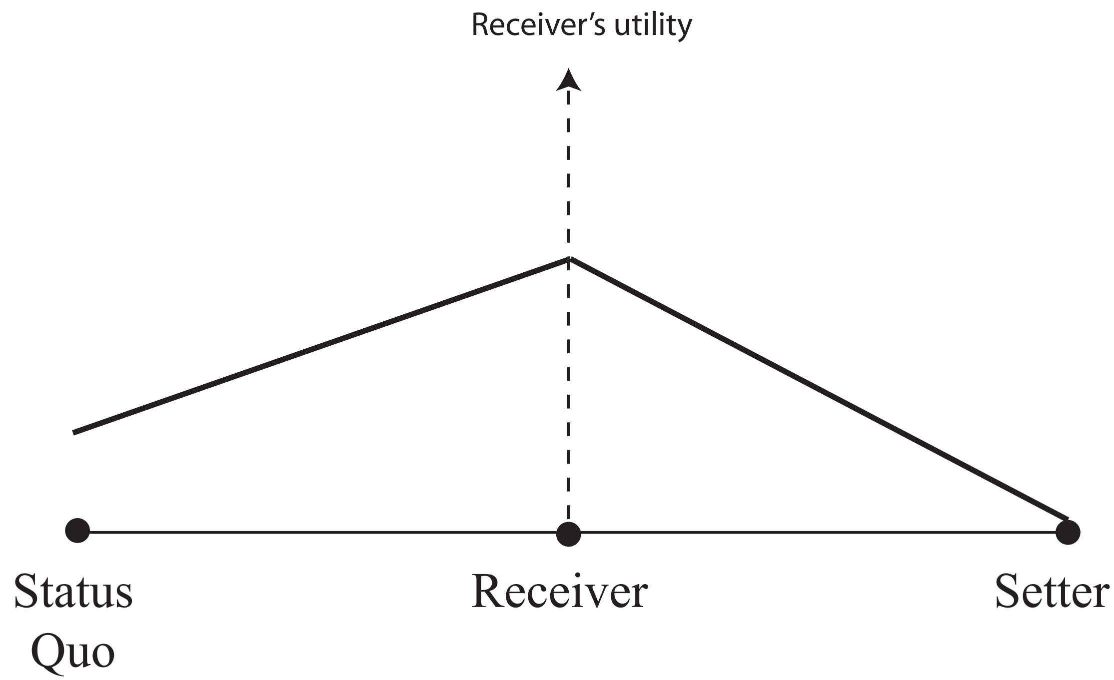
- Suppose that instead of choosing a physical ballot, voters must choose a representative, or agent, (with preferences of his own) to engage in a decision-making process with other representatives
- Voter has preferences $w \succ x \succ y \succ z \succ ...$
- Voter chooses an agent, and agent's preferences are aggregated with the preferences of other agents via some institutional rule, f, in order to generate an outcome
- Gibbard-Satterthwaite says that unless one agent is a dictator, there is no reason to expect that it is in the best interest of the voter to choose an agent whose preferences mirror his own

The Romer-Rosenthal setter model

A very commonly used, simple model of the institutional benefits accorded to an agenda setter

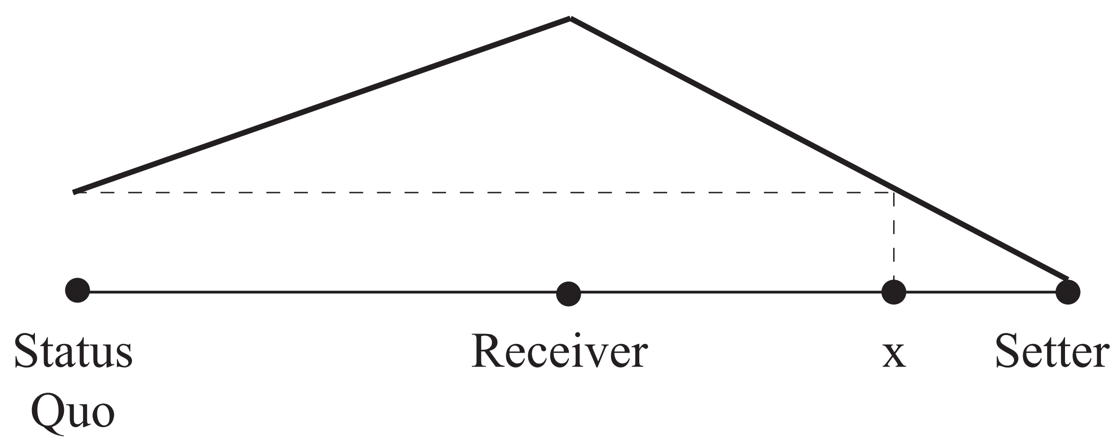
- There is an exogenous status quo, q
- An agenda setter proposes a policy change to the status quo, and if the receiver accepts it, it's implemented
- The setter chooses the policy that makes himself best off, subject to the constraint that the receiver needs to like it better than the status quo

The Romer-Rosenthal Setter Model

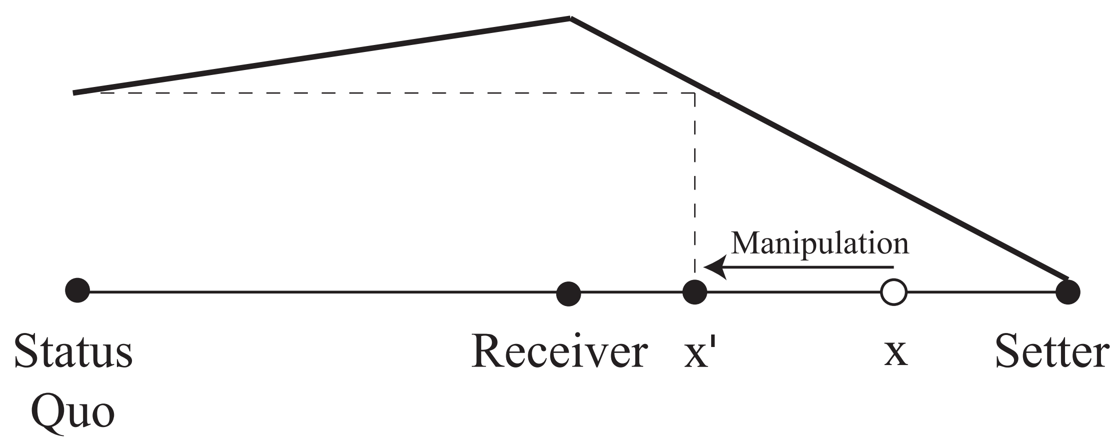


What policy will the setter propose?

The Romer-Rosenthal Setter Model



The Romer-Rosenthal Setter Model



Why this result is important

Misrepresentation of one's true preferences is not necessarily bad. But it is problematic for any problem relating to inference, and G-S says we should be careful when deducing preferences from behavior.

- How does a legislator's roll call vote reflect his or her policy preferences?
- How does an individual's vote choice reflect his or her preferences over the parties and/or candidates?
- Does the preference composition of a committee reflect the preferences of a group's members?
- Does an executive's best choice of political appointee necessarily pursue the policies that the executive would individually pursue in that position?

In-class question

		Voter 2 preferences	
		ABC	ACB
Voter 1 preferences	ABC	**ABC**	**ABC**
	CBA	**ABC**	**BCA**

1. Given the information in this table, can we conclude that our aggregation rule F satisfies *no dictator*?
2. Can we conclude that F satisfies *weak Pareto*?
3. Can we conclude that F satisfies *IIA*?
4. Now think about the above table as simply producing a single societal choice, which is the top-ranked alternative produced by F. Thus, at profile $\rho = (ABC, ABC)$, our societal choice is A. Given the information provided, is our choice function *strategy-proof*?

Possible lab questions

- How might we measure the “manipulability” of a voting rule?
- Given that successful manipulation requires some knowledge of the preferences of others, how could a definition of manipulability take varying degrees of this kind of knowledge into account?
- What are possible heuristics people could use in order to manipulate outcomes? How susceptible are different voting systems to these different heuristics?