Weak Strategyproofness in Randomized Social Choice

Felix Brandt
(joint work with Patrick Lederer, UNSW)

AAAI 2025, Philadelphia, USA
February 2025

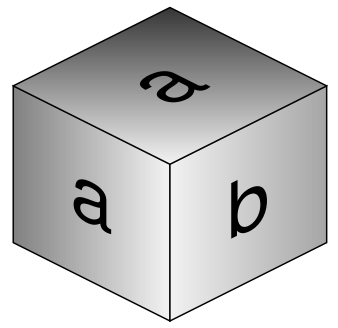
DSS
Decision Sciences & Systems

Randomized Social Choice

- N : Set of n voters
- $\succsim_i$: Preference relation of voter i over the m alternatives in A.
- $\mathscr{R}$: Set of all weak preference relations (complete, transitive)
- $\mathscr{L}$: Set of all strict preference relations (complete, transitive, and antisymmetric)
- $f\colon \mathscr{R}^N \to \Delta(A)$: Social decision scheme (SDS) maps $R = (\succsim_1, \ldots, \succsim_n)$ to a lottery $p \in \Delta(A)$.

a	b	a
b	a	c
c	c	b

Social Decision Scheme



$p(a) = 2/3$

$p(b) = 1/3$

- An SDS is
  - even-chance if it only returns uniform lotteries,
  - Condorcet-consistent if it puts probability 1 on Condorcet winners,
  - ex post efficient if it puts probability 0 on Pareto-dominated alternatives, and
  - strategyproof if no voter is better off by misstating his true preferences.

Stochastic Dominance

- When is a voter "better off"?
  - We only know his preferences over A, not his preferences over $\Delta(A)$.
- Quantify over all utility functions $u : A \to \mathbb{R}$ consistent with the voter's preference relation $\succsim$.
  - u is consistent with $\succsim$ iff $\forall x, y \in A\colon u(x) \geq u(y) \Leftrightarrow x \succsim y$.
- For $p, q \in \Delta(A)$,

$$p \succsim q \quad \text{iff } \forall u \in \mathbb{R}^A \text{ consistent with } \succsim\colon \mathbb{E}_p[u] \geq \mathbb{E}_q[u]$$

$$\text{iff } \forall x \in A\colon \textstyle\sum_{y \succsim x} p(y) \geq \sum_{y \succsim x} q(y).$$

- Some lotteries are incomparable ($\succsim$ is incomplete).

	a	b	c
	2	1	0
	7	5	1
	3	2	0
	$a >$	$b >$	c
$p=$ (	½	0	½)
$q=$ (	0	½	½)

	a	b	c
	4	3	0
	4	1	0
	$a >$	$b >$	c
$p=$ (	½	0	½)
$q=$ (	0	1	0)

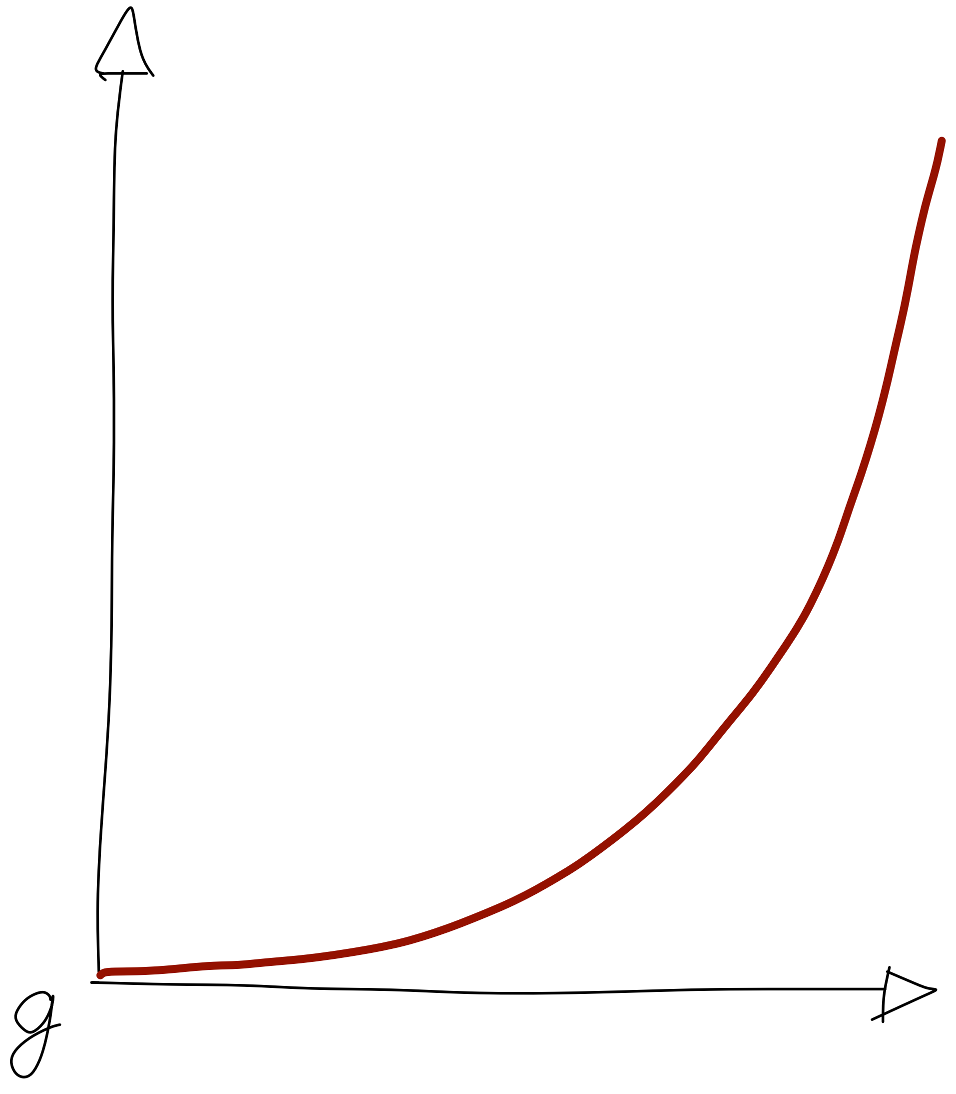
Strategyproofness

- The following has to hold for all R and $i \in N$.
- Strong strategyproofness: $\boxed{\forall \succsim_i' : \quad f(\succsim_i, \dots) \succsim_i f(\succsim_i', \dots)}$
    - A manipulation by i is successful if $\mathbb{E}_{f(\succsim_i', \dots)}[u_i] > \mathbb{E}_{f(\succsim_i, \dots)}[u_i]$ for ***some*** consistent $u_i \in \mathbb{R}^A$.
    - Gibbard (1977) gave a complete characterization of strongly strategyproof SDSs for $\mathscr{L}^N$.
- Weak strategyproofness: $\boxed{\forall \succsim_i' : \quad f(\succsim_i, \dots) \not\prec_i f(\succsim_i', \dots)}$
    - A manipulation by i is successful if $\mathbb{E}_{f(\succsim_i', \dots)}[u_i] > \mathbb{E}_{f(\succsim_i, \dots)}[u_i]$ for ***all*** consistent $u_i \in \mathbb{R}^A$.
    - Postlewaite & Schmeidler (SCW 1986); Bogomolnaia & Moulin (JET 2001)
    - Few SDSs were known to only satisfy weak strategyproofness:
        - Condorcet rule for $\mathscr{L}^N$ (Postlewaite & Schmeidler, SCW 1986)
        - Egalitarian simultaneous reservation for $\mathscr{R}^N$ (Aziz & Stursberg, AAAI 2014)
        - Omni* for $\mathscr{L}^N$ (Lederer, IJCAI 2021)

Results

Score Functions

- A function $s\colon \mathscr{L}^N \times A \to \mathbb{R}_{\geq 0}$ is a score function if for all R, $\sum_{x\in A} s(R,x) > 0$ and for all distinct $x, y, z \in A$ and $R' = R$, except that voter i swapped x and y such that $y \succ_i' x$.
  - $s(R,z) = s(R',z)$, (localizedness)
  - $s(R,y) \leq s(R',y)$, and (monotonicity)
  - $s(R,y) < s(R',y) \;\Rightarrow\; s(R,x) > s(R',x)$. (balancedness)
- Examples
  - Plurality: $s_P(R,x) = |\{i \in N\colon \forall y \in A\colon x \succsim_i y\}|$
  - Borda: $s_B(R,x) = \sum_{i\in N} |\{y \in A\colon x \succ_i y\}|$
  - Copeland: $s_C(R,x) = |\{y \in A\colon x \succ_{\text{maj}} y\}| + \frac{1}{2}|\{y \in A\setminus\{x\}\colon x \sim_{\text{maj}} y\}|$
- Let s and t be score functions and $g\colon \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ a strictly increasing function, then $s + t$ and $g \circ s$ are also score functions.
  - Boosted plurality, Borda, and Copeland scores: s_P^k, s_B^k, and s_C^k



Positive Results

- For score function s, the corresponding score-based SDS returns p with $p(x) = \frac{s(R,x)}{\sum_{y\in A} s(R,y)}$ for all $x \in A$.
- Theorem: Every score-based SDS for $\mathscr{L}^N$ is weakly strategyproof.
  - s_P^k, s_B^k, and s_C^k are arbitrarily good approximations of plurality, Borda, and Copeland.
    - Impossible with strong strategyproofness (Procaccia, AAAI 2010)!
    - These SDSs become manipulable for more and more utility functions as k increases.
- We can also allow infinite scores for at most one alternative.
  - E.g., Condorcet-consistent variant of s_C^k, which is approximately ex post efficient!
    - Impossible with strong strategyproofness (B. et al., SCW 2024)!
- We give a complete characterization of weakly strategyproof even-chance SDSs for $\mathscr{L}^N$ that only depend on the voters' top choices and are anonymous and neutral.

Limitations & Conclusion

- The following properties are incompatible with weak strategyproofness:
  1. **even-chance**, **Condorcet-consistency**, **ex post efficiency** $\mathscr{L}^N$, $m \geq 5$, $n \geq 5$ odd
     - open whether even-chance is required
  2. **pairwiseness**, **neutrality**, **ex post efficiency** $\mathscr{L}^*$, $m \geq 5$
  3. **anonymity**, **neutrality**, **ex ante efficiency** $\mathscr{R}^N$, $m \geq 4$, $n \geq 4$
     - much simpler proof than Brandl et al. (JACM 2018), 14 pages → 2 pages
     - still open whether neutrality is required
  4. **no bi-dictatorship**, **even-chance**, **ex post efficiency** $\mathscr{R}^N$, $m \geq 3$, $n \geq 3$
     - stronger than Corollary 2 of B. et al. (JET 2022)
     - are all ex post efficient, weakly strategyproof SDSs mixtures of dictatorships?
- We have identified a large class of interesting, weakly strategyproof SDSs.
- Several interesting questions concerning weak strategyproofness remain.