CS 405/6001: Game Theory and Algorithmic Mechanism Design **Jul–Nov 2026**

Lecture 16: Gibbard–Satterthwaite and Single-Peaked Preferences

Lecturer: Swaprava Nath *Scribe(s): Claude Scribe*

Disclaimer: *These notes aggregate content from several texts and have not been subjected to the usual scrutiny deserved by formal publications. If you find errors, please bring to the notice of the Instructor.*

16.1 Recap

Last lecture we set up social choice functions $f : \mathcal{P}^n \to A$, saw why Arrow's theorem pushes us toward studying SCFs rather than full social welfare functions, and proved that strategyproofness (SP) and monotonicity (MONO) are equivalent, which in turn forces PE, UN, and ONTO to coincide for any SP rule. Today we ask the sharper question those results were building toward: for $|A| \geqslant 3$, is there *any* SCF that is strategyproof and onto without being a dictatorship? The answer is the Gibbard–Satterthwaite theorem, and after proving it we look at how restricting the domain of preferences reopens the door to good rules.

16.2 The Gibbard–Satterthwaite Theorem

Theorem 16.1 (Gibbard 1973, Satterthwaite 1975) *Suppose $|A| \geqslant 3$. Then f is ONTO and SP if and only if f is dictatorial. Equivalently (by Corollary 15.26 of Lecture 15), the same holds with PE or UN in place of ONTO.*

Remark 16.2 (What this means) *No reasonable voting rule – plurality, Borda, Copeland, maximin, or any other onto rule – is strategyproof once there are three or more alternatives. This is a genuine impossibility, not a defect of any particular rule: it is a structural fact about the unrestricted domain $\mathcal{P}$, where every one of the $m!$ rankings is a feasible report and a would-be manipulator always has room to lie.*

Remark 16.3 (Points to note before the proof)

1. *At $|A| = 2$, Theorem 16.1* fails*: plurality with a fixed tie-break is SP, ONTO, and non-dictatorial.*
2. *The domain is all of $\mathcal{P}$. If preferences are restricted to a strict subset $\mathcal{S} \subset \mathcal{P}$, the theorem may no longer hold – this is exactly the escape route we pursue in Section 16.4.*
3. ***Indifference**: the theorem generally fails once ties in preference are permitted, unless the specific constructions used below remain possible under indifference.*
4. ***Cardinalization**: the theorem still holds under cardinal utilities, as long as every ordinal ranking remains achievable by some cardinal profile.*

16.3 Proof of the Gibbard–Satterthwaite Theorem

We follow a direct approach (Sen, 2001), proving the theorem first for $n = 2$ agents and then extending to $n \geqslant 3$ by induction on the number of agents.

Lemma 16.4 *Suppose $|A| \geqslant 3$, $N = \{1, 2\}$, and f is ONTO and SP. Then for every profile P, $f(P) \in \{P_1(1), P_2(1)\}$: the outcome is always one of the two agents' top choices.*

Proof: Case 1: $P_1(1) = P_2(1)$. Unanimity (equivalent to ONTO under SP, by Corollary 15.26) forces $f(P) = P_1(1)$ directly.

Case 2: $P_1(1) = a \neq b = P_2(1)$. Suppose for contradiction $f(P) = c \neq a, b$ (this needs $|A| \geqslant 3$). Construct P_2' with $a \succ_2' b$ but otherwise arbitrary, and P_1' with $b \succ_1' a$ but otherwise arbitrary, so that:

P_1	P_2	P_1	P_2'	P_1'	P_2'	P_1'	P_2
a	b	a	b	a	b	a	b
$\vdots$	$\vdots$	$\vdots$	a	b	a	b	$\vdots$

In (P_1, P_2'), every alternative except b is Pareto dominated by a (since P_1 ranks a first and P_2' ranks a above everything but possibly b), so $f(P_1, P_2') \in \{a, b\}$ by PE (equivalent to SP+ONTO). If $f(P_1, P_2') = b$, agent 2 profitably manipulates from true preference P_2 (which gives outcome c) to P_2' (which gives $b \succ_2 c$) – contradicting SP. So $f(P_1, P_2') = a$. By a symmetric argument, $f(P_1', P_2) = b$.

Now apply monotonicity. Going from (P_1, P_2') to (P_1', P_2'), alternative a's relative position is unchanged (it was already dominated only by things below b), so MONO forces $f(P_1', P_2') = a$. Going from (P_1', P_2) to (P_1', P_2'), alternative b's relative position is unchanged, so MONO forces $f(P_1', P_2') = b$. Since $a \neq b$, this is a contradiction. Hence $f(P) \in \{a, b\}$ after all. ■

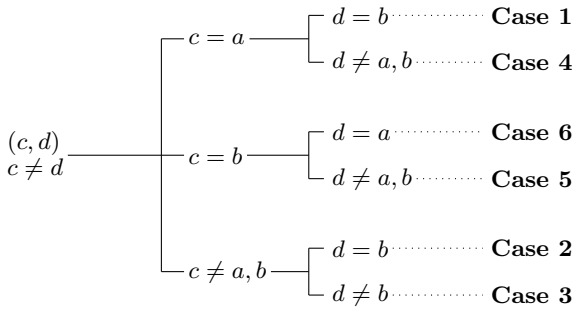
Lemma 16.5 (Two-player Gibbard–Satterthwaite) *Suppose $|A| \geqslant 3$, $N = \{1, 2\}$, f is ONTO and SP. Let $P : P_1(1) = a \neq b = P_2(1)$ and $P' : P_1'(1) = c, P_2'(1) = d$. If $f(P) = a$, then $f(P') = c$; if $f(P) = b$, then $f(P') = d$.*

This lemma says: once agent 1's top choice wins at some profile where the two agents disagree, agent 1's top choice wins at *every* such profile – agent 1 is a dictator (or symmetrically, agent 2 is).

Proof: If $c = d$, unanimity settles the claim immediately. So assume $c \neq d$. By symmetry it suffices to show $f(P) = a \implies f(P') = c$; six exhaustive cases on (c, d) cover every possibility.

case	c	d
1	a	b
2	$\neq a, b$	b
3	$\neq a, b$	$\neq b$
4	a	$\neq a, b$
5	b	$\neq a, b$
6	b	a

The same split is easier to check as a decision tree: branch first on c, then on d.



The tree also explains why there are six cases rather than nine. The profile with $c = d$ is already settled by unanimity, which removes $d = a$ from the $c = a$ branch and $d = b$ from the $c = b$ branch, leaving two leaves each. In the third branch, $d = a$ and $d \neq a, b$ are handled by the same argument, so they are merged into the single Case 3 condition $d \neq b$; note that stating Case 3 as $d \neq a, b$ instead would leave the pair ($c \neq a, b$, $d = a$) uncovered.

Case 1 ($c = a, d = b$): By Lemma 16.4, $f(P') \in \{a, b\}$. Suppose $f(P') = b$. Construct $\hat{P}$ with $\hat{P}_1 = a \succ b \succ \cdots$, $\hat{P}_2 = b \succ a \succ \cdots$. Since a's position is unchanged going from (P_1, P_2) to $(\hat{P}_1, \hat{P}_2)$, MONO gives $f(\hat{P}) = a$. Since b's position is unchanged going from (P'_1, P'_2) to $(\hat{P}_1, \hat{P}_2)$, MONO gives $f(\hat{P}) = b$. Contradiction, so $f(P') = a$.

Case 2 ($c \neq a, b$, $d = b$): By Lemma 16.4, $f(P') \in \{c, b\}$. Suppose $f(P') = b$. The transition $(P'_1, P'_2) \to (\hat{P}_1, P_2)$ (with $\hat{P}_1 = c \succ a \succ \cdots$) follows Case 1's pattern, giving $f(\hat{P}_1, P_2) = b$. But then agent 1, whose true preference is P_1 (with a on top), can manipulate from $\hat{P}_1$ to P_1 and flip the outcome to a, which they strictly prefer – contradicting SP. So $f(P') = c$.

Case 3 ($c \neq a, b$, $d \neq b$): By Lemma 16.4, $f(P') \in \{c, d\}$. Suppose $f(P') = d$. The transition $(P'_1, P'_2) \to (\hat{P}_1, \hat{P}_2)$ (with $\hat{P}_1 = c \succ \cdots$, $\hat{P}_2 = b \succ \cdots$) is an instance of Case 2 for agent 2, giving $f(\hat{P}) = b$. The transition $(P_1, P_2) \to (\hat{P}_1, \hat{P}_2)$ is an instance of Case 2 for agent 1, giving $f(\hat{P}) = c$. Since $b \neq c$, contradiction. So $f(P') = c$.

Case 4 ($c = a$, $d \neq a, b$): By Lemma 16.4, $f(P') \in \{a, d\}$. Suppose $f(P') = d$. As in Case 3, $(P'_1, P'_2) \to (\hat{P}_1, \hat{P}_2)$ is Case 2 for agent 2, giving $f(\hat{P}) = b$; and $(P_1, P_2) \to (\hat{P}_1, \hat{P}_2)$ is Case 1, giving $f(\hat{P}) = a$. Contradiction ($a \neq b$). So $f(P') = a$.

Case 5 ($c = b$, $d \neq a, b$): By Lemma 16.4, $f(P') \in \{b, d\}$. Suppose $f(P') = d$. The transition $(P'_1, P'_2) \to (\hat{P}_1, \hat{P}_2)$ is Case 1, giving $f(\hat{P}) = d$. Separately, $(P_1, P_2) \to (P_1, \hat{P}_2)$ is Case 4 for agent 1, giving $f(P_1, \hat{P}_2) = a$; and $(P_1, \hat{P}_2) \to (\hat{P}_1, \hat{P}_2)$ is Case 2 for agent 1, giving $f(\hat{P}) = b$. Now $f(\hat{P})$ cannot equal both d and b – contradiction. So $f(P') = b$.

Case 6 ($c = b$, $d = a$): Let $e \neq a, b$ and assume for contradiction $f(P') = a$. The transition $(P'_1, P'_2) \to (\hat{P}_1, \hat{P}_2)$ (with $\hat{P}_1 = e \succ \cdots$, $\hat{P}_2 = b \succ \cdots$) is Case 5 for agent 2, giving $f(\hat{P}) = b$. The transition $(P_1, P_2) \to (\hat{P}_1, \hat{P}_2)$ is Case 2 for agent 1, giving $f(\hat{P}) = e$. Since $e \neq b$, contradiction. So $f(P') = b$.

All six cases proved, so agent 1 is a dictator whenever the two agents' top choices differ; combined with Case 1 of Lemma 16.4 this makes agent 1 (or symmetrically agent 2) a dictator on all of $\mathcal{P}^2$. ∎

Remark 16.6 (Induction to $n \geqslant 3$) *The $n = 2$ case, proved above, is the base case. For $n \geqslant 3$, one shows the theorem holds for all $K < N$ and then extends to N by considering a reduced SCF g in which two agents have "coalesced" into one, leaving $N - 1$ effective agents; one shows g is SP and hence has a dictator by the inductive hypothesis, then argues this dictator induces a dictator in the original f. The full induction is in Sen (2001).*

16.4 Domain Restrictions and the Single-Peaked Domain

Theorem 16.1 is a statement about $f : \mathcal{P}^n \to A$, where $\mathcal{P}$ contains *every* strict ordering of A. As Remark 16.3 noted, this unrestricted domain is doing real work in the proof: a manipulator always has some misreport available. If we shrink the feasible preferences from $\mathcal{P}$ to a subset $\mathcal{S} \subset \mathcal{P}$, any f that was SP on $\mathcal{P}$ remains SP on $\mathcal{S}$, but new strategyproof rules may now exist that were not strategyproof (or not even well-defined as non-dictatorial options) on the full domain.

Example 16.7 (Three domain restrictions) *Three restrictions are classically studied:* ***single-peaked***

preferences *(alternatives lie on a line, each agent's ranking rises to a peak and falls on both sides),* ***divisible goods allocation*** *(preferences over shares of a divisible resource), and* ***quasi-linear preferences*** *(utility linear in money, enabling payments). We focus on the first.*

Definition 16.8 (Single-peaked preference) *Let $<$ be a common linear order on A (think of the real line). Preference P_i is* ***single-peaked*** *with respect to $<$ if, writing $P_i(1)$ for agent i's peak:*

1. *(left monotonicity) $\forall\, b, c \in A$ with $b < c \leqslant P_i(1)$: $c \succ_i b$;*
2. *(right monotonicity) $\forall\, b, c \in A$ with $P_i(1) \leqslant b < c$: $b \succ_i c$.*

Example 16.9 (GPU cluster power cap) *Several research teams share one multi-tenant GPU training cluster whose power cap is set once, cluster-wide. Each team i has an ideal cap t_i^*: too low throttles training, too high risks thermal shutdowns and higher electricity cost. Team i's satisfaction falls off the further the shared cap sits from t_i^* in* either *direction – exactly Definition 16.8 with the power-cap axis as the common order $<$. Other classic instances of the same structure: facility location (siting a school or hospital along a line) and political ideology (a left–center–right spectrum).*

Remark 16.10 (How much of a restriction is this?) *With m alternatives there are $m!$ strict orderings in the unrestricted domain $\mathcal{P}$, but only 2^{m-1} single-peaked orderings with respect to a fixed common order $<$. For example, with three power-cap settings low $<$ med $<$ high, all $3! = 6$ orderings exist, but only the $2^{3-1} = 4$ that move monotonically outward from a peak are single-peaked; the two orderings low $\succ$ high $\succ$ med and its mirror high $\succ$ low $\succ$ med are excluded, since satisfaction would have to rise, fall, then rise again.*

Proposition 16.11 (Leftmost-peak rule is strategyproof) *Let $\mathcal{S}$ be the single-peaked preferences with respect to $<$, and let $f : \mathcal{S}^n \to A$ select the leftmost reported peak. Then f is strategyproof, and so is the analogous rightmost-peak rule and, more generally, the rule selecting the fixed k-th peak from the left for any k.*

Proof: The team with the leftmost peak already receives its top choice under f, so it has no incentive to misreport. Any other team can only change the outcome by reporting a peak strictly further left than the current winner – but by single-peakedness (Definition 16.8), such an outcome is strictly worse for that team than the current winner, since it moves the outcome further from their true peak. The same argument, symmetrized, applies to the rightmost-peak rule and any fixed k-th-from-left rule, including the median ($k = \lceil n/2 \rceil$), which we study next. ∎

16.5 The Median Voter Theorem

Definition 16.12 (Median voter SCF) *$f : \mathcal{S}^n \to A$ is a* ***median voter SCF*** *if there exist fixed* ***phantom peaks*** *$B = \{y_1, \ldots, y_{n-1}\}$, independent of the profile P, such that*

$$f(P) = \operatorname{median}\big(B, \operatorname{peaks}(P)\big)$$

for every $P \in \mathcal{S}^n$, where the median is taken with respect to $<$.

If all $n-1$ phantoms sit at the leftmost feasible value, f coincides with the leftmost-peak rule; if all sit at the rightmost, f coincides with the rightmost-peak rule. Varying the phantom placements sweeps out the entire family of median voter SCFs.

Theorem 16.13 (Moulin, 1980) *Every median voter SCF is strategyproof.*

Proof: Fix a profile P with outcome $f(P) = a$, and suppose team i, whose true peak is b, considers misreporting a peak c, potentially moving the outcome to a'.

- If $a > b$ and $a > c$: c lies on the same side of a as b does, so the count of peaks (real and phantom) on each side of a is unchanged; $a' = a$.
- If $a > b$ and $a < c$: moving from b to c increases the count of peaks to the right of a by one, so the median shifts right, $a < a'$. By single-peakedness, team i (whose true peak $b < a$) strictly prefers a to a', so this does not benefit i.
- The two cases with $a < b$ are symmetric to the two above.

In every case, misreporting either leaves the outcome unchanged or strictly hurts team i, so no team benefits from lying. ■

Proposition 16.14 (PE characterization of median voter SCFs) *Let $p_{\min}, p_{\max}$ be the leftmost and rightmost reported peaks at profile P. A median voter SCF f satisfies $f(P) \in [p_{\min}, p_{\max}]$ if and only if f is Pareto efficient at P.*

Proof: ($\Longrightarrow$) Suppose $f(P) < p_{\min}$. Every team's peak lies at or to the right of $p_{\min} > f(P)$, so by single-peakedness every team prefers $p_{\min}$ to $f(P)$ – meaning $f(P)$ is Pareto dominated. Symmetric argument if $f(P) > p_{\max}$. Contrapositive: PE forces $f(P) \in [p_{\min}, p_{\max}]$.

($\Longleftarrow$) Suppose $f(P) \in [p_{\min}, p_{\max}]$ but some a Pareto dominates $f(P)$, i.e., $a \succ_i f(P)$ for every team i. If $a < f(P)$, the team at $p_{\max} \geqslant f(P)$ prefers $f(P)$ to a by single-peakedness (both lie on the same side of, or at, its peak) – contradiction. Symmetric if $a > f(P)$. So no such a exists, and $f(P)$ is PE. ■

Theorem 16.15 *Let $f : \mathcal{S}^n \to A$ be strategyproof. Then ONTO $\Longleftrightarrow$ UN $\Longleftrightarrow$ PE.*

Proof: PE $\Longrightarrow$ UN $\Longrightarrow$ ONTO holds as in the unrestricted domain. For the converse, suppose f is SP and ONTO but not PE: $\exists\, a, b$ with $a \succ_i b\ \forall i$, yet $f(P) = b$ at some profile P. Single-peakedness gives a neighbour c of b (possibly $c = a$) with $c \succ_i b\ \forall i$. Since f is ONTO, some P' has $f(P') = c$. Construct P'' with every team's peak at c, second choice b. Then $D(b, P_i) \subseteq D(b, P_i'')\ \forall i$ forces $f(P'') = b$ by monotonicity (equivalent to SP), while $D(c, P_i') \subseteq D(c, P_i'')\ \forall i$ forces $f(P'') = c$. Contradiction, so ONTO $\Longrightarrow$ PE under SP. ■

Definition 16.16 (Anonymity) *SCF f is **anonymous (ANON)** if for every profile P and every permutation σ of the agents, $f(P^\sigma) = f(P)$, where P^σ relabels agent i's preference to position $\sigma(i)$.*

Example 16.17 *With three teams $N = \{1, 2, 3\}$ and the cyclic permutation $\sigma(1) = 2, \sigma(2) = 3, \sigma(3) = 1$, the profile $P = (P_1, P_2, P_3)$ becomes $P^\sigma = (P_1^\sigma, P_2^\sigma, P_3^\sigma)$ with $P^\sigma_{\sigma(i)} = P_i$ for each i – that is, $P_1^\sigma = P_3$, $P_2^\sigma = P_1$, $P_3^\sigma = P_2$, a cyclic relabeling of which team supplied which column. Anonymity requires f to output the same alternative on P and P^σ.*

Remark 16.18 *Dictatorships are not anonymous: relabeling which agent is the dictator changes whose report determines the outcome, so permuting agent identities generally changes f's value.*

Theorem 16.19 (Median Voter Theorem, Moulin 1980) *A strategyproof SCF $f : \mathcal{S}^n \to A$ is ONTO and anonymous if and only if it is a median voter SCF.*

Proof:[Proof, "if" direction] Median voter SCFs are SP by Theorem 16.13. They are ANON: permuting agent identities does not change the *multiset* of reported peaks, hence does not change the median. They are ONTO: to reach any target a, have every agent report peak a; among the $2n-1$ points (the n agent peaks, all at a, and the $n-1$ fixed phantoms), the median is a regardless of where the phantoms sit. ■

Proof:[Proof sketch, "only if" direction] Given f SP, ANON, ONTO, we *construct* phantom peaks that make f agree with the median voter rule. Define P_i^0 to be the single-peaked preference with peak at the leftmost point of $<$, and P_i^1 with peak at the rightmost. Define

$$y_j = f\big(\underbrace{P_1^0, \ldots, P_{n-j}^0}_{n-j \text{ leftmost}}, \underbrace{P_{n-j+1}^1, \ldots, P_n^1}_{j \text{ rightmost}}\big), \qquad j = 1, \ldots, n-1.$$

By anonymity, only the count j matters, not which specific agents supply which extreme report, so this is well-defined. Set $B = \{y_1, \ldots, y_{n-1}\}$.

Claim 16.20 *$y_j \leqslant y_{j+1}$ for $j = 1, \ldots, n-2$.*

Proof: Comparing the profiles defining y_j and y_{j+1}, only agent $n-j$'s report differs: P_{n-j}^0 (peak leftmost) in the former, P_{n-j}^1 (peak rightmost) in the latter. By SP, agent $n-j$ cannot benefit from reporting P_{n-j}^1 when their true preference is P_{n-j}^0, so $y_j \succsim_{P_{n-j}^0} y_{j+1}$ or they are equal; since P_{n-j}^0 has its peak at the leftmost point, single-peakedness forces $y_j \leqslant y_{j+1}$. ■

With B constructed and ordered (Claim 16.20), one shows by a further monotonicity argument, agent by agent, that for every profile P with peaks $p_1 \leqslant \cdots \leqslant p_n$ (WLOG by ANON), $f(P) = \text{median}(p_1, \ldots, p_n, y_1, \ldots, y_{n-1})$ – i.e., f is exactly the median voter SCF defined by B. The case split (whether the median lands on a phantom or an agent peak) and the full monotonicity chase are given in Moulin (1980), "On strategy-proofness and single peakedness." ■

16.6 Summary and Preview

We proved the Gibbard–Satterthwaite theorem (Theorem 16.1): for $|A| \geqslant 3$, ONTO and SP together force dictatorship, first directly for $n = 2$ via Lemmas 16.4–16.5 and their six-case analysis, then by induction for $n \geqslant 3$. We then restricted the domain to single-peaked preferences (Definition 16.8, motivated by Example 16.9) and found that leftmost-, rightmost-, and median-peak rules are all strategyproof (Proposition 16.11). Median voter SCFs (Definition 16.12) are strategyproof by construction (Theorem 16.13), Pareto efficient exactly on $[p_{\min}, p_{\max}]$ (Proposition 16.14), and – together with ONTO and anonymity – give a complete characterization of the non-dictatorial strategyproof rules on the single-peaked domain (Theorem 16.19).

Next week: mechanism design with payments, where quasi-linear utilities let us design truthful, non-dictatorial rules even on the unrestricted preference domain.

References

[Gib73] A. Gibbard, "Manipulation of voting schemes: a general result," *Econometrica*, 41(4), 1973, pp. 587–601.

[Moulin80] H. Moulin, "On strategy-proofness and single peakedness," *Public Choice*, 35(4), 1980, pp. 437–455.

[Sat75] M. Satterthwaite, "Strategy-proofness and Arrow's conditions," *Journal of Economic Theory*, 10(2), 1975, pp. 187–217.

[Sen01] A. Sen, "Another direct proof of the Gibbard–Satterthwaite theorem," *Economics Letters*, 70(3), 2001, pp. 381–385.