Preference aggregation and impossibility

Robert J. Carroll

Last revised: 6 May 2026

1 Motivation

A democratic legislature considers three reform proposals; each member ranks them according to her own preferences; the legislature must collectively select one. A multi-member committee with conflicting preferences must produce a coherent set of recommendations, ranked from most to least supported. A constitution-drafter, designing institutions for an ideologically diverse polity, must specify aggregation rules that turn citizens' preferences into collective decisions. A voting reformer, choosing among plurality, Borda, instant-runoff, and approval voting, must understand which structural properties each rule does and does not satisfy. In each setting, the substantive question is the same: given a profile of individual preferences, what is the right rule for aggregating them into a collective preference or choice?

The welfare-economics framework of #28 left this question open. The Pareto criterion delivers a partial answer (a Pareto-improving alternative is unambiguously better than a Pareto-dominated one) but typically fails to determine a unique social choice; the welfare-function characterization theorem produces a Pareto optimum for any choice of weights, but the question of *which* weights to choose was not addressed by the theory. Bargaining theory (#29) supplies one possible answer in the two-player setting (the Nash bargaining solution corresponds to particular weights), but the question of how a society aggregates many individuals' preferences into a coherent collective preference is genuinely distinct from bilateral bargaining. Social choice theory — the substantive sub-field that takes up the aggregation question directly — is what we now develop.

The headline result is Arrow (1951)'s impossibility theorem: no rule for aggregating individual preferences into a social ordering simultaneously satisfies a small set of intuitive structural axioms (universal domain, weak Pareto, independence of irrelevant alternatives, non-dictatorship). The theorem is one of the most-cited results in twentieth-century social science, and its substantive reading has organized half a century of debate about democratic theory and welfare economics. Three structural insights organize this handout. First, the impossibility is robust: it persists when the axioms are weakened or modified in various ways (Wilson (1972) shows the impossibility holds without Pareto; Sen (1970b) shows a different impossibility involving minimal liberal rights). Second, the impossibility lifts to the social-choice-function level: Gibbard (1973) and Satterthwaite (1975) show that no social choice function with universal domain is simultaneously strategy-proof and non-dictatorial, and Muller and Satterthwaite (1977) show that the GS theorem is essentially the Arrow theorem in different language. Third, the substantive lessons of the impossibility cluster are not that aggregation is hopeless but that the structural shape of preference aggregation forces every aggregation rule to give up something — and the question of *what* to give up is the methodological choice that organizes the rest of social-choice theory (and the next handout).

This handout opens the social-choice cluster. The next handout takes up positive results under domain restrictions: Black (1948)'s median voter theorem under single-peaked preferences, the multi-dimensional chaos results of Plott (1967) and McKelvey (1976), single-crossing preferences, and related domain-restriction theorems. The forthcoming game-theory cluster will develop the strategic

/ mechanism-design side, where Vickrey–Clarke–Groves mechanisms and Myerson's optimal-auction theory live.

2 Social welfare functions and the aggregation problem

The basic objects are individuals, alternatives, and preferences. A working political-economy analyst tracking a collective decision typically has three primitive lists in hand: who the deciders are, what they're deciding among, and what each of them prefers. Aggregation theory asks what coherent collective ranking, or coherent collective choice, can be derived from these three primitives via a structurally well-behaved rule.

Throughout the handout we adopt the following standard conventions. There is a finite set $N = \{1, 2, \ldots, n\}$ of individuals (with $n \geq 2$) and a set X of alternatives (with $|X| \geq 3$ unless otherwise stated; the case $|X| = 2$ is special, as we'll see in §6).[1] Each individual $i \in N$ has a weak preference $\succsim_i$ on X, modeled as a complete and transitive binary relation. Let $\mathcal{R}$ denote the set of all complete-and-transitive binary relations on X (*weak orderings* or *preference relations*). A *preference profile* is an n-tuple $\boldsymbol{R} = (\succsim_1, \ldots, \succsim_n) \in \mathcal{R}^n$ recording each individual's preference. Strict preference is denoted $\succ_i$ (defined as $\succsim_i$ but not $\precsim_i$) and indifference $\sim_i$.

Definition 1 (Social welfare function and social choice function)**.** A *social welfare function* (SWF) is a map $F : \mathcal{R}^n \to \mathcal{R}$ that assigns to each preference profile a social preference relation $F(\boldsymbol{R}) \in \mathcal{R}$. We write $\succsim_F$ for the social preference, $\succ_F$ for the social strict preference, and $\sim_F$ for social indifference.

A *social choice function* (SCF) is a map $f : \mathcal{R}^n \to X$ (or, more generally, $f : \mathcal{R}^n \to 2^X \setminus \{\emptyset\}$) that assigns to each preference profile a chosen alternative (or non-empty set of chosen alternatives).

The distinction matters. An SWF produces a full ordering of all alternatives, including ties; an SCF produces just a winner (or a set of winners). Voting rules in actual elections are SCFs (we want a single winner, or a manageable set of finalists); committee deliberation procedures sometimes resemble SWFs (we want a ranked list of recommendations); welfare-economic analyses depend on SWFs (we want to know what's socially desirable across all alternatives, not just which is best). The aggregation theorems of §3–§5 are about SWFs; the strategy-proofness theorem of §6 is about SCFs. The Muller–Satterthwaite theorem connects them.

Example 2 (Three voting rules as SCFs)**.** Let $X = \{a, b, c\}$ and $N = \{1, 2, 3\}$.

- *Plurality rule*: each individual "votes" for her top-ranked alternative; the SCF picks the alternative with the most top-ranked votes (ties broken by some rule).
- *Borda count*: each individual gives $|X| - k$ Borda points to her k-th-ranked alternative ($|X| - 1$ to her top, $|X| - 2$ to her second, ..., 0 to her bottom); the SCF picks the alternative with the highest total Borda score.

[1] The $|X| \geq 3$ condition is essential for Arrow's theorem: the case $|X| = 2$ is structurally special, since on two alternatives every preference is a complete ranking and majority rule is well-defined and well-behaved (May (1952)'s theorem characterizes majority rule axiomatically on two-alternative settings). The structural reason Arrow's impossibility kicks in at $|X| \geq 3$ is that with three or more alternatives, the standard pairwise-majority comparisons can produce cycles (Condorcet cycles), and any aggregation rule that respects pairwise comparisons must address the resulting incoherence somehow — the IIA axiom is one specific way of forcing the issue.

- *Condorcet rule*: the SCF picks the alternative that majority-defeats every other in pairwise comparison (the *Condorcet winner*), if one exists; otherwise the rule is undefined or extended in some way.

Each of these rules is a function from preference profiles to alternatives, and each has different structural properties. The substantive point of the impossibility theorems is that no rule simultaneously satisfies all the structural properties one would intuitively want.

The classic Condorcet cycle illustrates one structural failure. Suppose voter 1's preference is $a \succ_1 b \succ_1 c$, voter 2's is $b \succ_2 c \succ_2 a$, voter 3's is $c \succ_3 a \succ_3 b$. In pairwise majority comparisons: a beats b ($1, 3$ vs. 2); b beats c ($1, 2$ vs. 3); c beats a ($2, 3$ vs. 1). The pairwise-majority relation is cyclic and therefore not transitive — there is no Condorcet winner. The aggregation question becomes: given that pairwise majority rule fails to produce a coherent ordering, what *can* we do? Arrow's theorem is the structural answer.

3 Arrow's theorem

Arrow's theorem isolates four structural axioms that a reasonable preference-aggregation rule should arguably satisfy and proves them collectively impossible (with $|X| \geq 3$). The methodology — characterize a class of rules by axioms, then show the axioms force impossibility — is the methodological foundation of the entire impossibility-theorem cluster, and it is the same methodological move at work in axiomatic bargaining (#29) and axiomatic decision theory (#26 / #27). Each axiom has a substantive reading; what makes the theorem powerful is that the axioms are individually defensible and collectively incoherent.

Definition 3 (Arrow's axioms)**.** A social welfare function $F : \mathcal{R}^n \to \mathcal{R}$ satisfies:

- **Universal domain (U):** F is defined on all profiles in $\mathcal{R}^n$ (no a-priori restriction on individual preferences).
- **Weak Pareto (P):** for every profile $\boldsymbol{R}$ and every pair $x, y \in X$, if $x \succ_i y$ for every i, then $x \succ_F y$.
- **Independence of irrelevant alternatives (IIA):** for every pair of profiles $\boldsymbol{R}, \boldsymbol{R}' \in \mathcal{R}^n$ and every pair $x, y \in X$, if individuals' rankings of x vs. y agree under $\boldsymbol{R}$ and $\boldsymbol{R}'$ (i.e., $x \succsim_i y \iff x \succsim_i' y$ for every i), then the social ranking of x vs. y agrees under $F(\boldsymbol{R})$ and $F(\boldsymbol{R}')$.
- **Non-dictatorship (D):** there is no $i \in N$ such that for every profile $\boldsymbol{R}$ and every pair $x, y \in X$ with $x \succ_i y$, $x \succ_F y$.

The substantive readings. (U) says the rule must work on every conceivable preference profile, not just the "nice" ones — a robustness condition. (P) says unanimity is respected: if everyone strictly prefers x to y, society must too. (IIA) says the social ranking of any pair depends only on individuals' rankings of *that* pair, not on rankings involving other alternatives. (D) says no single individual gets to determine the social ranking by herself — there is no dictator.

Theorem 4 (Arrow 1951)**.** *Let $|X| \geq 3$. There is no social welfare function $F : \mathcal{R}^n \to \mathcal{R}$ satisfying (U), (P), (IIA), and (D) simultaneously.*

Proof sketch. The standard proof is the *decisive-coalition* argument. Call a coalition $C \subseteq N$ *decisive* over the ordered pair (x, y) if for every profile in which everyone in C has $x \succ_i y$, the social preference is $x \succ_F y$. The proof proceeds in three steps.

Step 1 (Weak Pareto ⇒ N is decisive over every pair). Trivial: weak Pareto says exactly that N is decisive over every pair.

Step 2 (Field expansion). If a coalition C is decisive over some ordered pair (x, y) with $|X| \geq 3$, then C is decisive over every ordered pair (i.e., C is *globally decisive*). The proof uses (IIA) and (P) to extend decisiveness from one pair to others by considering profiles in which the relevant rankings are configured to force conclusions.

Step 3 (Group contraction). If a coalition C with $|C| \geq 2$ is globally decisive, then some strict subset $C' \subsetneq C$ is globally decisive. The proof partitions C into two non-empty subsets and uses (IIA) and (P) to show one of the subsets must be globally decisive over some pair, hence (by Step 2) globally decisive everywhere.

Iterating Step 3 starting from the globally decisive coalition N (Step 1), we conclude that some single individual is globally decisive — i.e., a dictator, violating (D). Austen-Smith and Banks (1999, Ch. 3) works through the proof in detail; Mas-Colell, Whinston, and Green (1995, Ch. 21) gives a complementary presentation. □

The decisive-coalition technique is one of the most powerful tools in social-choice theory and shows up repeatedly — in Wilson (1972), in Muller and Satterthwaite (1977), and elsewhere. The structural insight is that (P) plus (IIA) combine to make "decisiveness" a contagious property: any coalition that is decisive over some pair becomes decisive over every pair, and any coalition that is decisive at all must (if larger than a singleton) contain a strictly smaller decisive subcoalition. The combination forces decisiveness to "contract" all the way down to a single individual.

Example 5 (Three reform proposals)**.** A three-person commission considers three reform proposals: status quo s, moderate reform m, ambitious reform a. Suppose the preferences are:

Voter 1	Voter 2	Voter 3
a	m	s
m	s	a
s	a	m

Pairwise majority: a beats m $(1, 3$ vs. $2)$; m beats s $(1, 2$ vs. $3)$; s beats a $(2, 3$ vs. $1)$. A Condorcet cycle: pairwise majority rule fails IIA-respecting transitivity. By Arrow's theorem, any aggregation rule respecting unanimity, IIA, and no-dictatorship will likewise fail somewhere; the cycle is the simplest concrete diagnostic of the impossibility.

The substantive content of the theorem is not that aggregation is hopeless; it is that any aggregation rule with universal domain must give up at least one of (P), (IIA), or (D). Different rules give up different axioms: plurality and Borda rule give up IIA (the social ranking of a vs b can change when c is added or removed); supermajority rule (e.g., requiring $> 2/3$ to overturn the status quo) gives up (P) at the margins where supermajority fails; and the only rule that gives up only (D) is dictatorship itself. The methodological lesson is that the modeler choosing an aggregation rule is not choosing among "all the nice rules" but among rules that violate at least one structural property she would prefer to preserve — and the long-running debate about how to read this constraint substantively has been one of the central preoccupations of twentieth-century democratic theory.[2]

4 Sen's Paretian liberal

Arrow's theorem combines weak Pareto with IIA and non-dictatorship; Sen (1970b) shows that an even sparser combination — weak Pareto plus a minimal liberal-rights condition — is also impossible under universal domain. The result is conceptually distinct from Arrow's: Sen does not invoke IIA, and the impossibility shows that even a minimal commitment to individual liberty creates structural tension with Pareto efficiency. The substantive readings of Sen's paradox have been highly influential in liberal political philosophy and in the welfare-economics literature.

Definition 6 (Minimal liberalism)**.** A social welfare function F satisfies *minimal liberalism (ML)* if there exist at least two distinct individuals $i, j \in N$ and at least two distinct pairs $(x_i, y_i), (x_j, y_j)$ of alternatives in X such that:

- For every profile, the social preference between x_i and y_i agrees with i's preference: if $x_i \succ_i y_i$ then $x_i \succ_F y_i$, and conversely (and analogously for indifference);
- Similarly for j over (x_j, y_j).

The substantive reading: each of (at least) two individuals has a "personal sphere" — a pair of alternatives that differ only along some dimension that is privately hers, and her preferences over that dimension determine the social preference. Examples: the choice of what an individual reads in her personal time, what church she attends, what color she paints her bedroom. ML formalizes the claim that society respects each individual's preferences over her own personal sphere.

Theorem 7 (Sen 1970)**.** *There is no social welfare function $F : \mathcal{R}^n \to \mathcal{R}$ satisfying (U), (P), and (ML) simultaneously.*

Proof sketch. The standard demonstration is the "Lady Chatterley" example. Two individuals, prudish A and curious B, with three alternatives: "A reads it" (a), "B reads it" (b), and "no one reads it" (n). Preferences: A prefers $n \succ_A a \succ_A b$ (no one reads, but if someone must, A would rather it be him than B); B prefers $a \succ_B n \succ_B b$ (most preferred is A reading it, then no one, then B herself). Now apply the axioms:

[2] The substantive interpretation of Arrow's theorem has been the subject of an enormous philosophical-and-political-theory literature. Riker (1982) reads Arrow as a fundamental challenge to "populist" theories of democracy that take collective preferences as a coherent input to legitimacy: if the operation of going from individual preferences to collective preferences is structurally incoherent under intuitive axioms, then "what the people want" is not a well-defined object, and democratic legitimacy must rest on something other than coherent expression of popular will. Riker's preferred "liberal" reading takes Arrow as an argument for procedural / electoral democracy with constitutional constraints rather than for substantive popular sovereignty. Sen (1970a, 2002) reads Arrow more constructively: the theorem characterizes the structural shape of the aggregation problem and tells us which axiomatic compromises are necessary, not which ones are best; the methodological consequence is to study rules that satisfy specific subsets of the axioms (which is exactly what the next handout does, with single-peakedness and median voter theory). Coleman (1990) and the rational-choice tradition more generally read Arrow as a constraint that institutions must navigate rather than a fundamental impossibility: real institutions impose domain restrictions (single-peakedness, single-crossing, etc.) and accept axiomatic compromises (e.g., agenda-setter institutions that violate IIA) in ways that produce workable collective decisions.

- ML for A over (a, n): A has personal liberty over whether A reads, so $n \succ_F a$ (since $n \succ_A a$).
- ML for B over (b, n): B has personal liberty over whether B reads, so $n \succ_F b$ (since $n \succ_B b$).
- Pareto over (a, n): both prefer $a \succ n$? No, A has $n \succ a$. But over (a, b): both have $a \succ b$? Yes: A prefers a to b, and B prefers a to b. So Pareto gives $a \succ_F b$.
- Combined: $n \succ_F a$ (from ML), $a \succ_F b$ (from Pareto), $n \succ_F b$ (from ML), but also we need a coherent ranking, and we have $n \succ a \succ b$ by transitivity — consistent. But is this consistent with $b \succ n$? No, by ML we got $n \succ b$, so the ranking is $n \succ a \succ b$.

On this particular profile, no contradiction. The contradiction arises from a different profile: take A's preference as $b \succ_A n \succ_A a$ (prudish A would rather B read alone than himself read; he'd most prefer no one read but ranks himself reading worst since it would corrupt him personally) and B's as $n \succ_B a \succ_B b$. Then ML for A over (a, n) gives $n \succ_F a$, ML for B over (b, n) gives $n \succ_F b$, but Pareto over (a, b) gives $b \succ_F a$ (both prefer b to a). Combined: $n \succ_F b \succ_F a$ and $n \succ_F a$ are consistent, but if we add a third alternative ... The standard formal proof constructs a profile where the combined axioms produce a cycle. Sen (1970b) works through it; the cleanest exposition is in Sen (1970a, Ch. 6). □

The substantive reading is striking: the Pareto criterion — the most uncontroversial efficiency criterion of the welfare-economics framework — is in structural tension with even minimal individual liberty. The lesson is not that liberty is incompatible with efficiency in any deep sense, but that any aggregation rule that respects both must impose constraints on the preferences it can aggregate (i.e., must give up universal domain). The forthcoming domain-restriction handout explores what restrictions buy back consistency.

5 Wilson's theorem

Arrow's theorem requires the Pareto axiom (P) as one of its four conditions. Wilson (1972) asks: what does the impossibility look like if we drop (P)? The answer is that the impossibility persists in a different form, but with a more permissive class of rules.

Definition 8 (Imposed and inverse-imposed rules)**.** A social welfare function F is *imposed* (or *null*) if for every pair $x, y \in X$, the social preference between x and y is the same on every profile (i.e., the social preference is constant in $\boldsymbol{R}$, regardless of individuals' preferences). The function is *inverse-dictatorial* if there exists $i \in N$ such that for every profile and every pair x, y, $x \succ_i y$ implies $y \succ_F x$ (the social preference is the reverse of i's preference).

Theorem 9 (Wilson 1972)**.** *Let $|X| \geq 3$. Any social welfare function satisfying (U) and (IIA) is dictatorial, inverse-dictatorial, or imposed.*

The structural reading: dropping (P) does not avoid the Arrow-style impossibility; it only widens the class of "trivial" rules to include (a) dictatorship, (b) the reverse of dictatorship, and (c) constant rules that ignore individual preferences entirely. None of these is plausibly an aggregation rule one would want. The methodological lesson is that Pareto is not the source of Arrow's impossibility — something else is, structurally, and Wilson's theorem is the cleanest statement of what that "something else" looks like.

The structural lesson of §3–§5 taken together: there is no escape from Arrow-style impossibility by tweaking the axioms. Whether we add (Sen) or drop (Wilson) the Pareto condition, whether we focus on full-ordering aggregation (Arrow, Sen, Wilson) or on weaker aggregation conditions, the impossibility is robust. The escape route, when there is one, is to give up universal domain — to restrict the class of profiles the rule must operate on — which is the methodological heart of the next handout.

6 Strategy-proofness and Gibbard–Satterthwaite

So far we have studied SWFs — rules that produce a full social ordering. The corresponding question for SCFs — rules that produce just a single chosen alternative — is whether there is an SCF that is *strategy-proof*, i.e., one for which no individual ever has an incentive to misreport her preferences. Gibbard (1973) and Satterthwaite (1975) independently established the negative answer: no non-trivial SCF with universal domain is strategy-proof. The result is the SCF-side analogue of Arrow's theorem, and Muller and Satterthwaite (1977) show that the two are essentially equivalent under a structural translation.

Definition 10 (Strategy-proofness)**.** A social choice function $f : \mathcal{R}^n \to X$ is *strategy-proof* if for every individual i, every profile $\boldsymbol{R} = (\succsim_1, \dots, \succsim_i, \dots, \succsim_n)$, and every alternative preference $\succsim_i'$,

$$f(\boldsymbol{R}) \succsim_i f(\boldsymbol{R}')$$

where $\boldsymbol{R}' = (\succsim_1, \dots, \succsim_i', \dots, \succsim_n)$ differs from $\boldsymbol{R}$ only in i's reported preference. Strategy-proofness says that truth-telling is a (weakly) dominant strategy for every individual: no individual ever benefits from misreporting.

Definition 11 (Dictatorial SCF)**.** An SCF f is *dictatorial* if there exists $i \in N$ such that for every profile $\boldsymbol{R}$, $f(\boldsymbol{R})$ is i's top-ranked alternative under $\succsim_i$.

Theorem 12 (Gibbard 1973, Satterthwaite 1975)**.** *Let $|X| \geq 3$, and let $f : \mathcal{R}^n \to X$ be a social choice function with full range (i.e., every alternative in X is selected by f on some profile). If f is strategy-proof, then f is dictatorial.*

The substantive reading is parallel to Arrow's: under universal domain and minimal richness on the alternative space, strategy-proofness forces dictatorship. The full-range condition replaces non-dictatorship as the substantive non-triviality requirement (without it, an SCF could just always pick a fixed alternative regardless of preferences, which is trivially strategy-proof but uninteresting).

Theorem 13 (Muller–Satterthwaite 1977)**.** *Under standard regularity conditions, a social choice function f is strategy-proof and non-dictatorial* if and only if *the corresponding social welfare function (constructed from f by aggregating the choice-function outputs across menus) satisfies Arrow's axioms (U), (P), (IIA), and (D). Equivalently: GS-style impossibility for SCFs is structurally the same impossibility as Arrow's for SWFs.*

The Muller–Satterthwaite link is conceptually important. It shows that the Gibbard–Satterthwaite theorem is not a different impossibility from Arrow's; it is the same impossibility expressed in different vocabulary. Strategy-proofness (a behavioral / strategic property of an SCF) corresponds to a particular structural property of the corresponding SWF (*monotonicity*), and the conjunction of strategy-proofness with full range and non-dictatorship maps onto Arrow's axiomatic conjunction.

The methodological lesson is that the impossibility of strategy-proof aggregation is not a separate phenomenon to be discovered by studying mechanism design; it is a consequence of the same structural fact about preference aggregation that Arrow's theorem identifies.

Example 14 (Plurality voting and strategic manipulation)**.** Three voters with $X = \{a, b, c\}$. Voter 1's true preference: $a \succ_1 b \succ_1 c$. Voter 2's: $b \succ_2 a \succ_2 c$. Voter 3's: $c \succ_3 a \succ_3 b$. Under plurality, voter 1 votes for a, voter 2 for b, voter 3 for c — a three-way tie. Suppose tie-breaking favors c (alphabetical reverse, say). Then voter 1, who prefers $a \succ b \succ c$, can manipulate by reporting her preference as $b \succ_1 a \succ_1 c$ (voting for b instead of a). Then b wins outright with 2 votes, and voter 1 gets her second-best outcome (b) instead of her worst (c). Plurality is not strategy-proof.

The substantive lesson: every commonly-used voting rule with universal domain is manipulable. The Gibbard–Satterthwaite theorem says that this is structurally inevitable, not a curable defect of particular rules. The escape routes are the same as for Arrow: restrict the domain (e.g., to single-peaked preferences, where median voting becomes strategy-proof), or accept manipulability and design rules that are manipulable in less consequential ways.

7 What's next

This handout opens the social-choice cluster (#30–#31). The next handout takes up positive results under domain restrictions: Black (1948)'s median voter theorem under single-peaked preferences, the Plott / McKelvey chaos results in multi-dimensional policy spaces, single-crossing preferences as a related domain restriction, and May (1952)'s axiomatic characterization of majority rule on two alternatives. The structural through-line is that universal-domain impossibility (this handout) gives way to existence-and-uniqueness of well-behaved rules under appropriate domain restrictions (next handout).

The forthcoming game-theory cluster will develop the strategic / mechanism-design side. The Gibbard–Satterthwaite theorem of §6 is the structural foundation; mechanism design takes up the question of what implementable choice functions look like under weaker assumptions (Bayesian rather than dominant-strategy implementation, monetary transfers, restricted preference domains). Vickrey (1961) and Clarke (1971) and Groves (1973) develop the VCG mechanism for quasi-linear environments where transfers are available; Myerson (1981) develops the optimal-auction characterization for Bayesian incentive-compatible mechanisms. Both lines of work are forward-pointers from the present handout's GS theorem.

For graduate-level treatments at this handout's level: Mas-Colell, Whinston, and Green (1995, Ch. 21) is the standard micro-theoretic reference; Austen-Smith and Banks (1999, 2005) are the canonical political-science-native references and treat both the impossibility and the positive theory in detail; Sen (1970a) is the classic philosophical treatment with extensive substantive discussion; Gaertner (2009) is the standard graduate textbook on social-choice theory.

8 Exercises

Exercise 15. *Condorcet cycle.* Construct a profile of three voters over three alternatives $\{a, b, c\}$ that produces a Condorcet cycle (i.e., a majority-beats b, b majority-beats c, c majority-beats a). Verify the cycle and discuss in two sentences why the cycle implies that pairwise majority rule is not a coherent SWF.

Exercise 16. *Plurality and IIA.* Three voters, alternatives $\{a, b, c\}$, preferences: voter 1 has $a \succ b \succ c$; voter 2 has $b \succ c \succ a$; voter 3 has $c \succ b \succ a$. (a) Find the plurality winner. (b) Now add a fourth alternative d to the menu, with all voters ranking d in some way that doesn't change their pairwise rankings of a, b, c (e.g., everyone ranks d at the top). What does plurality select now? (c) Use the example to argue that plurality violates IIA.

Exercise 17. *Borda and IIA.* Construct an example showing that the Borda count violates IIA. (Hint: a profile where introducing a new alternative changes the Borda ranking of two existing alternatives.)

Exercise 18. *Verifying Arrow's axioms.* Consider the SWF defined by "the social preference is the unanimous preference if it exists, and otherwise the social preference is voter 1's preference." (a) Verify this SWF satisfies (U). (b) Does it satisfy (P)? (IIA)? (D)? (c) Use Arrow's theorem to identify which axiom must be violated, and verify your answer.

Exercise 19. *Sen's Lady Chatterley.* Work through the Lady Chatterley example with a profile that produces an explicit cycle in the social preference under (P) + (ML). (a) Specify the alternatives and preferences. (b) Apply (ML) and (P) to derive social preferences. (c) Identify the cycle. (d) Discuss in two sentences the substantive lesson: what does the example say about the tension between Pareto efficiency and individual liberty?

Exercise 20. *Strategic manipulation under plurality.* Construct an example with three voters and three alternatives in which a voter has an incentive to strategically misreport her preferences under plurality. (a) Specify the true preferences. (b) Specify the strategic misreport and resulting plurality outcome. (c) Verify that the manipulating voter is better off under the misreport. (d) Discuss in two sentences how the example illustrates the Gibbard–Satterthwaite theorem.

Exercise 21. *Wilson's theorem applied.* Consider the SWF that always declares social indifference between every pair of alternatives, regardless of the profile. (a) Verify it satisfies (U) and (IIA). (b) Identify which of the three Wilson outcomes (dictatorial, inverse-dictatorial, imposed) it represents. (c) Discuss in one sentence why this SWF is not a useful aggregation rule, despite satisfying (U) and (IIA).

Exercise 22. *Constitutional convention.* A three-person commission designs an aggregation rule for collective decisions. They are committed to (U) (the rule must work on any preferences) and (D) (no dictator). (a) Argue that they must give up either (P) or (IIA). (b) Suppose they decide IIA is too strong (the social ranking of a vs b should respond to the introduction of c). What kind of aggregation rule could they then adopt? (c) Suppose instead they decide weak Pareto is the dispensable axiom. What kind of aggregation rule could they adopt then?

Exercise 23. *Agenda-setter institutions and IIA.* Consider an agenda-setter procedure: the agenda-setter chooses a fixed binary agenda (alternative a vs. alternative b) and the committee votes pairwise. (a) Argue that the agenda-setter procedure violates IIA at the institutional level: the social ranking of a vs. b depends on the agenda-setter's choice rather than only on individuals' rankings of a vs. b. (b) Discuss in two or three sentences how the violation of IIA is what makes agenda-setter institutions an interesting object of political-economy study (cf. McKelvey (1976) and Riker (1980) on agenda-setter power).

Exercise 24. *Substantive interpretation.* In two or three sentences, articulate the substantive significance of Arrow's theorem for democratic theory. Does the impossibility theorem imply that democracy is structurally incoherent, or that any specific institutional design must accept particular axiomatic compromises? Cite Riker (1982) on the populist-vs-liberal reading of Arrow.

References

Arrow, Kenneth J. (1951). *Social Choice and Individual Values.* New York: Wiley.

Austen-Smith, David and Jeffrey S. Banks (1999). *Positive Political Theory I: Collective Preference.* Ann Arbor: University of Michigan Press.

– (2005). *Positive Political Theory II: Strategy and Structure.* Ann Arbor: University of Michigan Press.

Black, Duncan (1948). "On the Rationale of Group Decision-making". In: *Journal of Political Economy* 56.1, pp. 23–34.

Clarke, Edward H. (1971). "Multipart Pricing of Public Goods". In: *Public Choice* 11, pp. 17–33.

Coleman, James S. (1990). *Foundations of Social Theory.* Cambridge, MA: Belknap Press of Harvard University Press.

Gaertner, Wulf (2009). *A Primer in Social Choice Theory.* revised. Oxford: Oxford University Press.

Gibbard, Allan (1973). "Manipulation of Voting Schemes: A General Result". In: *Econometrica* 41.4, pp. 587–601.

Groves, Theodore (1973). "Incentives in Teams". In: *Econometrica* 41.4, pp. 617–631.

Mas-Colell, Andreu, Michael D. Whinston, and Jerry R. Green (1995). *Microeconomic Theory.* New York: Oxford University Press.

May, Kenneth O. (1952). "A Set of Independent Necessary and Sufficient Conditions for Simple Majority Decision". In: *Econometrica* 20.4, pp. 680–684.

McKelvey, Richard D. (1976). "Intransitivities in Multidimensional Voting Models and Some Implications for Agenda Control". In: *Journal of Economic Theory* 12.3, pp. 472–482.

Muller, Eitan and Mark A. Satterthwaite (1977). "The Equivalence of Strong Positive Association and Strategy-proofness". In: *Journal of Economic Theory* 14.2, pp. 412–418.

Myerson, Roger B. (1981). "Optimal Auction Design". In: *Mathematics of Operations Research* 6.1, pp. 58–73.

Plott, Charles R. (1967). "A Notion of Equilibrium and Its Possibility under Majority Rule". In: *American Economic Review* 57.4, pp. 787–806.

Riker, William H. (1980). "Implications from the Disequilibrium of Majority Rule for the Study of Institutions". In: *American Political Science Review* 74.2, pp. 432–446.

– (1982). *Liberalism Against Populism: A Confrontation between the Theory of Democracy and the Theory of Social Choice.* San Francisco: W. H. Freeman.

Satterthwaite, Mark A. (1975). "Strategy-proofness and Arrow's Conditions: Existence and Correspondence Theorems for Voting Procedures and Social Welfare Functions". In: *Journal of Economic Theory* 10.2, pp. 187–217.

Sen, Amartya K. (1970a). *Collective Choice and Social Welfare.* San Francisco: Holden-Day.

– (1970b). "The Impossibility of a Paretian Liberal". In: *Journal of Political Economy* 78.1, pp. 152–157.

– (2002). *Rationality and Freedom.* Cambridge, MA: Harvard University Press.

Vickrey, William (1961). "Counterspeculation, Auctions, and Competitive Sealed Tenders". In: *Journal of Finance* 16.1, pp. 8–37.

Wilson, Robert (1972). "Social Choice Theory without the Pareto Principle". In: *Journal of Economic Theory* 5.3, pp. 478–486.