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

Lecture 14: Arrow's Impossibility Theorem

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.*

14.1 Recap and Motivation

Last lecture we studied mechanism design under the assumption that agents might lie about their private information, and we saw how the Revelation Principle lets us focus on direct, truthful mechanisms. Today we set that concern aside entirely. Suppose every agent reports their true preference, honestly, with no strategic behaviour at all. A more basic question remains: *can we always combine everyone's honest individual preferences into one clear, consistent social ranking?* This is the question of **social choice theory**, pioneered by Kenneth Arrow.

Example 14.1 (A concrete version) *A large open-source foundation's steering committee has several members, each ranking a slate of competing technical proposals by personal preference. Is there always a reasonable rule for merging the n individual rankings into one committee ranking?*

14.2 Preference Orderings and Social Welfare Functions

Definition 14.2 (Preference relation) *A finite set of alternatives $A = \{a_1, \ldots, a_m\}$, $|A| \geqslant 3$, and a finite set of agents $N = \{1, \ldots, n\}$. Each agent i has a preference relation R_i over A, where aR_ib means "i finds a at least as good as b."*

Definition 14.3 (Preference ordering) *R_i is an **ordering** if it is*

- ***Complete**: for any $a, b \in A$, either aR_ib or bR_ia (or both);*
- ***Transitive**: aR_ib and bR_ic together imply aR_ic.*

Let $\mathcal{R}$ denote the set of all such orderings over A. Any ordering R_i decomposes into a strict preference P_i (aP_ib: aR_ib holds but bR_ia does not) and an indifference relation I_i (aI_ib: both aR_ib and bR_ia hold).

Definition 14.4 (Arrovian Social Welfare Function (ASWF)) *An ASWF is a function $F : \mathcal{R}^n \to \mathcal{R}$ taking a profile of individual orderings $(R_1, \ldots, R_n)$ to a single* collective *ordering $F(R)$, itself complete and transitive. We write the strict part of $F(R)$ as $\hat{F}(R)$.*

Crucially, an ASWF must output a full ordering of *every* alternative, not merely select a single winner.

14.3 Pareto Properties

Definition 14.5 (Weak Pareto) *F satisfies Weak Pareto if, for any pair $a, b \in A$ and any profile R,*

$$[\forall i \in N,\ aP_i b] \implies [a\hat{F}(R)b].$$

Definition 14.6 (Strong Pareto) *F satisfies Strong Pareto if, for any pair $a, b \in A$ and any profile R,*

$$\big[(\forall i \in N,\ aR_i b) \text{ and } (\exists j \in N,\ aP_j b)\big] \implies [a\hat{F}(R)b].$$

Strong Pareto implies Weak Pareto, since the hypothesis of Weak Pareto (everyone strictly prefers a) is a special case of Strong Pareto's hypothesis (everyone weakly prefers a, someone strictly). Both are minimal fairness requirements: if every committee member ranks proposal A over proposal B, it would be perverse for the aggregate ranking to disagree.

14.4 Independence of Irrelevant Alternatives

Example 14.7 (Two agreeing profiles) *Consider four alternatives a, b, c, d under two profiles R and R':*

$$R = \begin{pmatrix} a & a & c & d \\ b & c & b & c \\ c & b & a & b \\ d & d & d & a \end{pmatrix} \qquad R' = \begin{pmatrix} d & c & b & b \\ a & a & c & a \\ b & b & a & d \\ c & d & d & c \end{pmatrix}$$

where each column lists one agent's ranking, top to bottom. For every agent, the relative order of a and b is unchanged between R and R': both profiles agree *on (a, b), even though c and d move around.*

Definition 14.8 (Agreement on a pair) *R_i and R'_i agree on (a, b) if $aP_i b \Leftrightarrow aP'_i b$, $bP_i a \Leftrightarrow bP'_i a$, and $aI_i b \Leftrightarrow aI'_i b$. Write $R_i|_{\{a,b\}} = R'_i|_{\{a,b\}}$; if this holds for every i, write $R|_{\{a,b\}} = R'|_{\{a,b\}}$.*

Definition 14.9 (Independence of Irrelevant Alternatives (IIA)) *F satisfies IIA if, whenever $R|_{\{a,b\}} = R'|_{\{a,b\}}$, then $F(R)|_{\{a,b\}} = F(R')|_{\{a,b\}}$.*

The social ranking of a versus b should depend only on how agents individually rank a against b, not on where the "irrelevant" alternatives c, d happen to sit in anyone's ranking. If the profiles disagree on (a, b), IIA is vacuously satisfied.

Remark 14.10 *A recommender system ranking two songs A and B for a user should not flip their relative order merely because an unrelated third song C was added to the catalogue. That flip is precisely an IIA violation, and it is a real failure mode of some recommendation algorithms.*

14.5 Scoring Rules Cannot Satisfy IIA

Scoring rules are the most natural large family of ASWFs, so before proving Arrow's general theorem it is worth seeing that this entire family already fails.

Definition 14.11 (Scoring rule) *Fix scores $s_1 \geqslant s_2 \geqslant \cdots \geqslant s_m$, one per rank position. Given a profile R, each agent awards s_p to the alternative they place in position p. The total score of alternative x is*

$$\sigma(x) \;=\; \sum_{i \in N} s_{\,\mathrm{pos}_i(x)},$$

where $\mathrm{pos}_i(x)$ is the position of x in agent i's ranking. The social ordering $F(R)$ ranks alternatives in weakly decreasing order of total score. For simplicity we take each R_i to be a strict ordering here, so positions are well defined.

Borda ($s = (m-1, m-2, \ldots, 0)$), plurality ($s = (1, 0, \ldots, 0)$) and veto ($s = (1, \ldots, 1, 0)$) are all of this form.

Remark 14.12 (Weak Pareto for scoring rules) *A scoring rule satisfies Weak Pareto if and only if $s_1 > s_m$. If $s_1 > s_m$ and every agent ranks a above b, then each agent contributes weakly more score to a than to b, and strictly more whenever the two occupy positions with distinct scores, so $\sigma(a) > \sigma(b)$. Conversely if $s_1 = s_m$ then all the scores coincide, every alternative receives ns_1 in every profile, and $F(R)$ declares everything tied, violating Weak Pareto.*

Before the main result, a point that is easy to get wrong.

Proposition 14.13 (A dictatorship is not a scoring rule) *For $n \geqslant 2$, no scoring rule is a dictatorship.*

Proof: Every scoring rule is **anonymous**: $\sigma(x)$ is a sum over agents, so permuting which agent submitted which ballot leaves every total score unchanged, hence leaves $F(R)$ unchanged. A dictatorship with dictator d is not anonymous when $n \geqslant 2$: take any profile in which agents d and $j \neq d$ have different orderings, and swap their two ballots. The dictatorship's output changes from R_d to R_j, while any anonymous rule's output cannot change. ■

So the family of scoring rules and the single rule "dictatorship" are disjoint for $n \geqslant 2$, and the failure of scoring rules to satisfy IIA has no dictatorial escape hatch inside that family. The correct statement is the following.

Theorem 14.14 (No scoring rule satisfies IIA) *Let $m \geqslant 3$ and $n \geqslant 2$. A scoring rule satisfies IIA if and only if $s_1 = s_m$, i.e. if and only if it is the trivial rule declaring all alternatives tied in every profile. Consequently, no scoring rule satisfies Weak Pareto and IIA together.*

The "if" direction is immediate: when $s_1 = s_m$ the output ordering is the same (all tied) for every profile, so IIA holds vacuously, and Weak Pareto fails by Remark 14.12. The content is the "only if" direction, which we prove by explicit construction.

14.5.1 The idea behind the construction

A scoring rule compares a and b only through the sign of $\sigma(a) - \sigma(b)$. Agent i's ballot contributes

$$s_{\,\mathrm{pos}_i(a)} - s_{\,\mathrm{pos}_i(b)}$$

to this difference. Here is the key observation: whether agent i ranks a above b fixes only the *sign* of this contribution, never its *magnitude*. The magnitude depends on how far apart a and b sit in i's ballot, and that spacing is controlled entirely by where the irrelevant alternatives are placed. So by shuffling only the

irrelevant alternatives, we can hold every agent's a-versus-b opinion fixed while re-weighting how loudly each agent's opinion counts. If the available magnitudes are not all equal, the aggregate verdict can be flipped.

Accordingly, define

$$\Delta_{\min} \;=\; \min_{p<q}(s_p - s_q), \qquad \Delta_{\max} \;=\; \max_{p<q}(s_p - s_q) \;=\; s_1 - s_m,$$

the smallest and largest possible magnitudes of a single agent's contribution.

Claim 14.15 *Write $d_j = s_j - s_{j+1} \geqslant 0$ for $j = 1, \dots, m-1$. Then $\Delta_{\min} = \min_j d_j$, and if $s_1 > s_m$ then $\Delta_{\max} \geqslant (m-1)\Delta_{\min}$, and in particular $\Delta_{\max} > \Delta_{\min}$.*

Proof: For $p < q$ we have $s_p - s_q = \sum_{j=p}^{q-1} d_j \geqslant \min_j d_j$, since every $d_j \geqslant 0$; and taking $q = p+1$ attains d_p. Hence $\Delta_{\min} = \min_j d_j$. Also $\Delta_{\max} = \sum_{j=1}^{m-1} d_j \geqslant (m-1)\min_j d_j = (m-1)\Delta_{\min}$, and $m \geqslant 3$ gives $\Delta_{\max} \geqslant 2\Delta_{\min}$. If $\Delta_{\min} = 0$ then $\Delta_{\max} = s_1 - s_m > 0 = \Delta_{\min}$; if $\Delta_{\min} > 0$ then $\Delta_{\max} \geqslant 2\Delta_{\min} > \Delta_{\min}$. ■

The strict gap $\Delta_{\max} > \Delta_{\min}$ is the entire engine of the construction.

14.5.2 The general construction

Proof of Theorem 14.14, "only if" direction: Assume $s_1 > s_m$. Let p_0 be an index attaining the minimum consecutive drop, so $s_{p_0} - s_{p_0+1} = \Delta_{\min}$ by Claim 14.15. Define two ballot shapes for an agent who ranks a above b:

- **narrow**: a in position p_0, b in position $p_0 + 1$, the remaining $m-2$ alternatives filling the other positions in any fixed order. Contribution to $\sigma(a) - \sigma(b)$: $+\Delta_{\min}$.
- **wide**: a in position 1, b in position m, the rest in between. Contribution: $+\Delta_{\max}$.

Mirror images (swapping the roles of a and b) serve agents who rank b above a, contributing $-\Delta_{\min}$ and $-\Delta_{\max}$ respectively.

Partition the agents into $k_1 \geqslant 1$ agents of *type 1*, who will rank a above b, and $k_2 \geqslant 1$ of *type 2*, who will rank b above a, with $k_1 + k_2 = n$. Build the two profiles:

	type 1 ballots	type 2 ballots
R	narrow	wide
R'	wide	narrow

Then

$$\sigma_R(a) - \sigma_R(b) = k_1\Delta_{\min} - k_2\Delta_{\max}, \qquad \sigma_{R'}(a) - \sigma_{R'}(b) = k_1\Delta_{\max} - k_2\Delta_{\min}.$$

Every type 1 agent ranks a above b in both profiles and every type 2 agent ranks b above a in both, so $R|_{\{a,b\}} = R'|_{\{a,b\}}$: the profiles agree on the pair, and only the irrelevant alternatives have moved.

It remains to choose k_1, k_2 making the two social verdicts on (a, b) differ. Subtracting the two displayed quantities,

$$\big(\sigma_{R'}(a) - \sigma_{R'}(b)\big) - \big(\sigma_R(a) - \sigma_R(b)\big) \;=\; (k_1 + k_2)\big(\Delta_{\max} - \Delta_{\min}\big) \;=\; n\big(\Delta_{\max} - \Delta_{\min}\big) \;>\; 0,$$

so the two differences are never equal; we need only place 0 between them, i.e. arrange

$$k_1\Delta_{\min} \leqslant k_2\Delta_{\max} \qquad \text{and} \qquad k_1\Delta_{\max} \geqslant k_2\Delta_{\min}.$$

If $\Delta_{\min} = 0$, both hold for any split with $k_1, k_2 \geqslant 1$. Otherwise set $r = \Delta_{\min}/\Delta_{\max}$, which lies in $(0, \frac{1}{2}]$ by Claim 14.15; the two conditions read $r \leqslant k_1/k_2 \leqslant 1/r$. For n even take $k_1 = k_2 = n/2$, giving ratio $1 \in [r, 1/r]$. For n odd with $n \geqslant 3$, take $k_1 = (n-1)/2$ and $k_2 = (n+1)/2$, giving ratio $(n-1)/(n+1) \geqslant \frac{1}{2} \geqslant r$ and $< 1 \leqslant 1/r$.

With such a split, $\sigma_R(a) - \sigma_R(b) \leqslant 0 \leqslant \sigma_{R'}(a) - \sigma_{R'}(b)$ with at least one inequality strict, so $F(R)|_{\{a,b\}} \neq F(R')|_{\{a,b\}}$ while $R|_{\{a,b\}} = R'|_{\{a,b\}}$. IIA is violated. ∎

14.5.3 A single counterexample that defeats every scoring rule

For $m = 3$ and $n = 3$ one fixed pair of profiles works for every scoring rule at once, with no case analysis on where the minimum drop occurs. This is the version used in lecture.

Example 14.16 (Three alternatives, three voters) *Let $s_1 \geqslant s_2 \geqslant s_3$ with $s_1 > s_3$ be arbitrary, and consider*

$$R = \begin{pmatrix} a & b & b \\ b & c & c \\ c & a & a \end{pmatrix} \qquad R' = \begin{pmatrix} a & b & c \\ c & a & b \\ b & c & a \end{pmatrix}$$

each column being one voter's ranking, top to bottom.

The profiles agree on (a, b)***.*** *Voter 1 ranks a above b in both ($a \succ b \succ c$ in R, $a \succ c \succ b$ in R'). Voters 2 and 3 rank b above a in both ($b \succ c \succ a$ in R; $b \succ a \succ c$ and $c \succ b \succ a$ in R'). Only the position of c changes. Hence $R|_{\{a,b\}} = R'|_{\{a,b\}}$.*

Scores in R***.***

$$\sigma(a) = s_1 + s_3 + s_3, \qquad \sigma(b) = s_2 + s_1 + s_1,$$

$$\sigma(b) - \sigma(a) = (2s_1 + s_2) - (s_1 + 2s_3) = (s_1 - s_3) + (s_2 - s_3) \; > \; 0,$$

strictly, because $s_1 > s_3$ and $s_2 \geqslant s_3$. So $F(R)$ ranks b strictly above a.

Scores in R'***.***

$$\sigma(a) = s_1 + s_2 + s_3, \qquad \sigma(b) = s_3 + s_1 + s_2,$$

so $\sigma(a) = \sigma(b)$ and $F(R')$ ranks a and b as tied.

Therefore $R|_{\{a,b\}} = R'|_{\{a,b\}}$ but $F(R)|_{\{a,b\}} \neq F(R')|_{\{a,b\}}$, for every choice of $s_1 \geqslant s_2 \geqslant s_3$ with $s_1 > s_3$. IIA fails.

Example 14.17 (Instantiations of Example 14.16) *Borda, $s = (2, 1, 0)$: in R, $\sigma(a) = 2$ and $\sigma(b) = 5$, so b wins the pair; in R', $\sigma(a) = \sigma(b) = 3$, a tie. Plurality, $s = (1, 0, 0)$: in R, $\sigma(a) = 1$ and $\sigma(b) = 2$; in R', $\sigma(a) = \sigma(b) = 1$. Veto, $s = (1, 1, 0)$: in R, $\sigma(a) = 1$ and $\sigma(b) = 3$; in R', $\sigma(a) = \sigma(b) = 2$.*

Remark 14.18 (Where the degenerate cases live) *Two boundary cases are worth keeping straight, since they are the source of a common confusion.*

- *$s_1 = s_m$, the constant rule: satisfies IIA, fails Weak Pareto. By Theorem 14.14 this is the* only *scoring rule satisfying IIA when $n \geqslant 2$.*
- *$n = 1$: a strictly decreasing scoring rule outputs $F(R) = R_1$, which satisfies IIA and is trivially a "dictatorship". This single-agent case is the only setting in which a scoring rule and a dictatorship coincide, and it falls outside the scope of Arrow's theorem.*

In particular it is not *correct to say that a dictatorship is the scoring rule escaping IIA: by Proposition 14.13 a dictatorship is not a scoring rule at all once $n \geqslant 2$.*

Theorem 14.14 rules out one large and natural family. Arrow's theorem, proved next, is far stronger: it rules out *every* ASWF, scoring or otherwise, leaving only dictatorships.

Definition 14.19 (Dictatorship) *An ASWF F is a dictatorship if there exists a fixed agent $d \in N$ such that $F(R) = R_d$ for every profile R.*

14.6 Arrow's Impossibility Theorem

Theorem 14.20 (Arrow, 1951) *For $|A| \geqslant 3$, if an ASWF F satisfies Weak Pareto and IIA, then F must be a dictatorship.*

This is not a statement that most rules fail; *every* ASWF satisfying these two mild-looking axioms is forced into a dictatorship. The proof identifies a **decisive** group of agents and then shrinks that group until only one member remains.

14.6.1 Decisiveness

Definition 14.21 (Almost decisive and decisive) *Let $F : \mathcal{R}^n \to \mathcal{R}$ be given, $G \subseteq N$, $G \neq \emptyset$.*

1. *G is **almost decisive** over $\{a, b\}$ if for every R:*
$$[aP_i b,\ \forall i \in G, \quad bP_j a,\ \forall j \in N \setminus G] \implies [a\hat{F}(R)b].$$
We write this with the shorthand $\overline{D}_G(a, b)$: "G is almost decisive over $\{a, b\}$ w.r.t. F".
2. *G is **decisive** over $\{a, b\}$ if for every R:*
$$[aP_i b,\ \forall i \in G] \implies [a\hat{F}(R)b].$$
We write this with the shorthand $D_G(a, b)$: "G is decisive over $\{a, b\}$ w.r.t. F".

Observation: $D_G(a, b) \implies \overline{D}_G(a, b)$, since the premise of almost decisiveness is a special case of the premise of decisiveness. The converse need not hold in general. Decisiveness is also not symmetric: $D_G(a, b)$ says nothing about $D_G(b, a)$.

Example 14.22 (Reading the two notions) *If a small clique on the steering committee satisfies $D_G(A, B)$, that clique's preference for proposal A over proposal B always wins, regardless of what the rest of the committee thinks. If it only satisfies $\overline{D}_G(A, B)$, the clique is guaranteed to win only in the extreme situation where everyone else unanimously prefers B to A.*

The proof of Theorem 14.20 proceeds in two parts.

- **Part 1** (Field expansion lemma): if a group is *almost decisive* over a pair of alternatives, it is *decisive* over all pairs of alternatives.
- **Part 2** (Group contraction lemma): if a group is decisive, then a strict non-empty subset of that group is also decisive.

These two lemmas immediately prove the theorem.

14.6.2 Field Expansion Lemma

Lemma 14.23 (Field Expansion Lemma, FEL) *Let F satisfy WP and IIA. Then $\forall a, b, x, y$, $G \subseteq N$, $G \neq \emptyset$, $a \neq b$, $x \neq y$:*

$$\overline{D}_G(a, b) \Rightarrow D_G(x, y).$$

It implies that under WP and IIA, the two notions of decisiveness are equivalent. The cases to consider, ordered for the convenience of the proof, are:

1. $\overline{D}_G(a, b) \Rightarrow D_G(a, y)$, $y \neq a, b$
2. $\overline{D}_G(a, b) \Rightarrow D_G(x, b)$, $x \neq a, b$
3. $\overline{D}_G(a, b) \Rightarrow D_G(x, y)$, $x \neq a, b$, $y \neq a, b$
4. $\overline{D}_G(a, b) \Rightarrow D_G(x, a)$, $x \neq a, b$
5. $\overline{D}_G(a, b) \Rightarrow D_G(b, y)$, $y \neq a, b$
6. $\overline{D}_G(a, b) \Rightarrow D_G(a, b)$
7. $\overline{D}_G(a, b) \Rightarrow D_G(b, a)$

Case 1: $\overline{D}_G(a, b) \Rightarrow D_G(a, y)$, $y \neq a, b$: Pick an arbitrary $R \in \mathcal{R}^n$ such that aP_iy, $\forall i \in G$. We need to show $a\hat{F}(R)y$. Construct R' as follows:

G		$N \setminus G$	
a	a	b	b
$\vdots$	$\vdots$	$\vdots$	$\vdots$
b	b	a	y
$\vdots$	$\vdots$	$\vdots$	$\vdots$
y	y	y	a

positions of a and y in $N \setminus G$ s.t. $R'|_{a,y} = R|_{a,y}$.

That is, agents in G rank a above b above y; agents outside G rank b on top, with a and y below b in whichever order they had them in R.

- $\overline{D}_G(a, b) \Rightarrow a\hat{F}(R')b$, since in R' every agent in G has aP_ib and every agent outside G has bP_ja.
- WP over $b, y \Rightarrow b\hat{F}(R')y$, since every agent ranks b above y in R'.
- Transitivity $\Rightarrow a\hat{F}(R')y$.
- IIA $\Rightarrow a\hat{F}(R)y$, since $R'|_{a,y} = R|_{a,y}$. Hence $D_G(a, y)$.

■

Case 2: $\overline{D}_G(a, b) \Rightarrow D_G(x, b)$, $x \neq a, b$: Pick an arbitrary $R \in \mathcal{R}^n$ such that xP_ib, $\forall i \in G$. We need to show $x\hat{F}(R)b$. Construct R' as follows:

G		$N \setminus G$	
x	x	x	b
$\vdots$	$\vdots$	$\vdots$	$\vdots$
a	a	b	x
$\vdots$	$\vdots$	$\vdots$	$\vdots$
b	b	a	a

positions of x and b in $N \setminus G$ s.t. $R'|_{x,b} = R|_{x,b}$.

That is, agents in G rank x above a above b; agents outside G rank a at the bottom, with x and b above a in whichever order they had them in R.

- $\overline{D}_G(a,b) \Rightarrow a\hat{F}(R')b$, since in R' every agent in G has aP_ib and every agent outside G has bP_ja.
- WP over $x, a \Rightarrow x\hat{F}(R')a$, since every agent ranks x above a in R'.
- Transitivity $\Rightarrow x\hat{F}(R')b$.
- IIA $\Rightarrow x\hat{F}(R)b$, since $R'|_{x,b} = R|_{x,b}$. Hence $D_G(x,b)$.

■

Other cases. The remaining cases follow by chaining Cases 1 and 2 with the observation $D_G \Rightarrow \overline{D}_G$ (marked "definition"):

- **Case 3:** $\overline{D}_G(a,b) \overset{\text{(case 1)}}{\Longrightarrow} D_G(a,y)\ (y \neq a,b) \overset{\text{(definition)}}{\Longrightarrow} \overline{D}_G(a,y) \overset{\text{(case 2)}}{\Longrightarrow} D_G(x,y)\ (x \neq a,y)$
- **Case 4:** $\overline{D}_G(a,b) \overset{\text{(case 2)}}{\Longrightarrow} D_G(x,b)\ (x \neq a,b) \overset{\text{(definition)}}{\Longrightarrow} \overline{D}_G(x,b) \overset{\text{(case 1)}}{\Longrightarrow} D_G(x,a)\ (x \neq a,b)$
- **Case 5:** $\overline{D}_G(a,b) \overset{\text{(case 1)}}{\Longrightarrow} D_G(a,y)\ (y \neq a,b) \overset{\text{(definition)}}{\Longrightarrow} \overline{D}_G(a,y) \overset{\text{(case 2)}}{\Longrightarrow} D_G(b,y)\ (y \neq a,b)$
- **Case 6:** $\overline{D}_G(a,b) \overset{\text{(case 2)}}{\Longrightarrow} D_G(x,b)\ (x \neq a,b) \overset{\text{(definition)}}{\Longrightarrow} \overline{D}_G(x,b) \overset{\text{(case 2)}}{\Longrightarrow} D_G(a,b)$
- **Case 7:** $\overline{D}_G(a,b) \overset{\text{(case 5)}}{\Longrightarrow} D_G(b,y)\ (y \neq a,b) \overset{\text{(definition)}}{\Longrightarrow} \overline{D}_G(b,y) \overset{\text{(case 1)}}{\Longrightarrow} D_G(b,a)$

This completes the proof of Lemma 14.23. ■

14.6.3 Group Contraction Lemma

Lemma 14.24 (Group Contraction Lemma, GCL) *Let F satisfy WP and IIA, and let $G \subseteq N$, $G \neq \emptyset$, $|G| \geqslant 2$ be decisive. Then $\exists G' \subset G$, $G' \neq \emptyset$ which is also decisive.*

Proof: G with $|G| \geqslant 2$ is given. Let $G_1 \subset G$ and $G_2 = G \setminus G_1$, with $G_1, G_2 \neq \emptyset$, arbitrary. Construct R:

G_1	G_2	$N \setminus G$
a	c	b
b	a	c
c	b	a

$aP_ib,\ \forall i \in G$ and G decisive $\Rightarrow a\hat{F}(R)b$.

Where can c stand in $F(R)$ with respect to a? We show that in every possible case, either G_1 or G_2 is decisive.

Case 1: $a\hat{F}(R)c$**.** Consider G_1. In R, aP_ic for all $i \in G_1$ and cP_ia for all $i \in N \setminus G_1$. Consider each R' in which this relation holds; each such R' agrees with R on $\{a,c\}$, so by IIA $a\hat{F}(R')c$. Hence $\overline{D}_{G_1}(a,c) \overset{\text{(FEL)}}{\Longrightarrow} D_{G_1}$.

Case 2: $\neg(a\hat{F}(R)c) \implies cF(R)a$**.** Since $F(R)$ is complete, failing to rank a strictly above c means c is ranked at least as high as a. Then $a\hat{F}(R)b$ and $cF(R)a$ give $c\hat{F}(R)b$ by transitivity. Consider G_2. In R, cP_ib for all $i \in G_2$ and bP_ic for all $i \in N \setminus G_2$. Consider each R' in which this relation holds; by IIA $c\hat{F}(R')b$. Hence $\overline{D}_{G_2}(c,b) \overset{\text{(FEL)}}{\Longrightarrow} D_{G_2}$.

In either case a strict non-empty subset of G is decisive. ■

14.6.4 Proof of Arrow's Theorem

Proof of Theorem 14.20: By WP, N is decisive: $D_N(a,b)$ for every pair, since if every agent prefers a to b, so does the social ranking. Apply the Group Contraction Lemma repeatedly: N decisive implies a strict non-empty subset is decisive, which implies a smaller one is decisive, and so on; since N is finite and each step strictly shrinks the set, this ends at a singleton $\{d\}$ that is decisive. Then

$$D_{\{d\}}(a,b) \overset{\text{(definition)}}{\Longrightarrow} \overline{D}_{\{d\}}(a,b) \overset{\text{(FEL)}}{\Longrightarrow} D_{\{d\}}(x,y) \quad \text{for every pair } (x,y).$$

So agent d's strict preference determines the social ranking on every pair of alternatives, which is exactly the definition of a dictatorship. ∎

Remark 14.25 *The uncomfortable takeaway: any aggregation rule satisfying two axioms that individually sound almost too weak to matter, unanimity (Weak Pareto) and independence from irrelevant alternatives, collapses entirely to one person's opinion. Real voting and ranking systems must give up at least one of Weak Pareto, IIA, an unrestricted domain of preferences, or transitivity of the aggregate output, in order to escape this conclusion. The rest of the course, starting next week, studies exactly these escape routes: restricted domains, weaker notions of strategy-proofness, and specific classes of social choice functions like the median voter rule.*

14.7 Summary and Preview

We introduced Arrovian Social Welfare Functions and two natural fairness axioms: Weak/Strong Pareto (unanimity should carry through to the social ranking) and Independence of Irrelevant Alternatives (the social ranking of a pair should depend only on individual rankings of that pair). We then proved that the entire family of scoring rules fails IIA (Theorem 14.14), by exhibiting two profiles that agree on a pair yet produce different social verdicts on it. The mechanism of the counterexample is that an agent's ranking of a against b fixes only the sign of that agent's contribution to the score gap, while the irrelevant alternatives control its magnitude. We were careful to note that a dictatorship is not itself a scoring rule once $n \geqslant 2$ (Proposition 14.13); the only scoring rule satisfying IIA is the trivial constant one, which fails Weak Pareto. Arrow's theorem then generalises this failure from scoring rules to *all* ASWFs: for at least three alternatives, Weak Pareto and IIA together force a dictatorship. The proof identified decisive sets of agents and used the Field Expansion Lemma (power over one pair spreads to power over all pairs) together with the Group Contraction Lemma (decisive groups can always be shrunk) to isolate a single dictator. Next week we study how to escape this impossibility: the Gibbard–Satterthwaite theorem, which restates the impossibility in the language of strategy-proofness, and domain restrictions such as single-peaked preferences, under which good aggregation rules again become possible.

References

[Arr51] K. J. Arrow, "Social Choice and Individual Values," Wiley, 1951.

[MSZ13] M. Maschler, E. Solan and S. Zamir, *Game Theory*, Chapter 10, Cambridge University Press, 2013.