MANIFOLD MATHEMATICS

MONOGRAPH SERIES · VOLUME 14

The Arithmetic of Disagreement

Rank Aggregation, Social Choice, and What Reciprocal Rank Fusion Is Really Doing

A self-contained course in the mathematics of combining ordered lists, from the symmetric group to hybrid retrieval

MANIFOLD MATHEMATICS · VISAKHAPATNAM · AUGUST 2026

Where this monograph sits

Primary subject. *Social choice theory* — specifically the theory of *rank aggregation* — studied through the *combinatorics of the symmetric group*.

Mathematics Subject Classification (MSC 2020).

91B12	Voting theory
91B14	Social choice theory; Arrow's theorem
05A05	Permutations, words, matrices
06A07	Combinatorics of partially ordered sets
62G30	Order statistics; nonparametric inference
68P20	Information storage and retrieval
68Q17	Computational difficulty of problems (NP-hardness of Kemeny)

The three parents. This subject is a genuine hybrid. It inherits *objects* from algebra and combinatorics (the symmetric group $\mathfrak{S}_n$, its metrics, its group actions), *questions* from mathematical economics (which aggregation rules are fair? which are possible at all?), and *urgency* from computer science (hybrid search, recommender systems, retrieval-augmented generation, ensemble learning).

Prerequisites. Elementary group theory (permutations, cycle decomposition, group actions), basic metric-space language, elementary probability, and comfort with Jensen's inequality. No measure theory. No topology. Chapters 9 and 10 assume familiarity with maximum likelihood but define everything used.

Examination relevance. Material bearing directly on CSIR-NET (Part C combinatorics and algebra), GATE MA (linear algebra, probability), and Olympiad-style combinatorics is flagged in the margin and summarised in Appendix C.

On the problems. There are 58 problems in three graded sets. Full solutions are supplied for Sets A and B; Set C carries detailed hints and, where a problem is genuinely open-ended, a discussion of what a good answer looks like.

Manifold Mathematics monographs are written to be read with a pen. Every computation in this volume was verified numerically; the verification scripts are described in Appendix B.

Contents

Preface: Three Friends and a Restaurant

This monograph grew out of a diagram that circulated on the internet. Three friends in Pune — call them Shailesh, Rohan and Ashish — wanted dinner. Three restaurants were on the table. Each friend wrote down a ranking. The diagram then showed how a formula borrowed from information retrieval, *Reciprocal Rank Fusion*, picks a winner, and drew a moral: the formula rewards the candidate everyone quietly agreed on, rather than the one somebody shouted loudest about.

The diagram is charming and the formula is real. The moral, as we shall prove in Chapter 5, is *false* — or rather, it is true of a different formula that happens to be hiding inside this one. The reciprocal is not what produces consensus. Something else in the procedure is, and identifying it correctly changes how you should deploy the method in practice.

That is the intellectual arc of this book. We will take a piece of engineering folklore, place it inside its proper mathematical home, and find that the folklore is a reasonable approximation to a true statement about a *neighbouring* object. Along the way we will need the symmetric group, four different metrics on it, the Condorcet paradox, Arrow's impossibility theorem, the NP-hardness of finding a median permutation, and two probabilistic models of ranked data.

The second restaurant problem. A reader of the original diagram posed a variant. Keep the three friends; rotate their ballots:

Shailesh:	Vaishali $\succ$ SP's $\succ$ Shabree
Rohan:	SP's $\succ$ Shabree $\succ$ Vaishali
Ashish:	Shabree $\succ$ Vaishali $\succ$ SP's

Every restaurant now receives exactly one first place, one second, one third. What does the formula say?

It says nothing. All three scores are identical to twelve decimal places. Pairwise majority voting does worse than nothing: it produces a cycle, Vaishali $\succ$ SP's $\succ$ Shabree $\succ$ Vaishali. The Kemeny median has three co-optimal answers.

This is not a failure of any particular method. Chapter 5 (Theorem 5.12) shows that *every* rule satisfying a single, utterly innocent fairness axiom must return a perfect tie on this profile. The tie is forced by group theory.

The book is organised in five parts. Part I places the subject. Part II builds the geometry of ranked data. Part III is the theory of aggregation rules, and contains the main results about RRF. Part IV covers probabilistic models and the retrieval application that motivated everything. Part V is problems.

A reader who wants only the punchline should read Chapter 5 and stop. A reader preparing for an examination should read Chapters 2, 3, 6, 7 and do Problem Set B. A reader building a retrieval system should read Chapters 5 and 10 and take the diagnostic in §10.6 seriously.

Part I

Placement

CHAPTER 1

What Subject Is This?

1.1 The question is not rhetorical

A student who meets Reciprocal Rank Fusion in a machine-learning course, and Arrow's theorem in an economics course, and the symmetric group in an algebra course, has met one subject three times without being told. This chapter names it.

> **Answer.** The subject is *rank aggregation*, a branch of *social choice theory*. Its objects are elements of the symmetric group $\mathfrak{S}_n$ and its generalisations; its questions are about maps
>
> $$\Phi : \mathfrak{S}_n^m \longrightarrow \mathfrak{S}_n \qquad \text{or} \qquad \Phi : \mathfrak{S}_n^m \to \mathbb{R}^n,$$
>
> and which properties such maps can and cannot simultaneously enjoy.

1.2 The genealogy

1.2.1 Root: Condorcet and Borda, 1770s

The subject is older than most of undergraduate mathematics. In 1770 Jean-Charles de Borda proposed to the Académie des Sciences that elections be decided by summing positions rather than counting first-place votes. In 1785 the Marquis de Condorcet published his *Essai*, containing the observation that majority preference can be cyclic — a discovery so counterintuitive that it was largely forgotten for a century and a half.

Both men were doing what we would now call combinatorics of permutations, without the vocabulary.

1.2.2 Trunk: Arrow, 1951

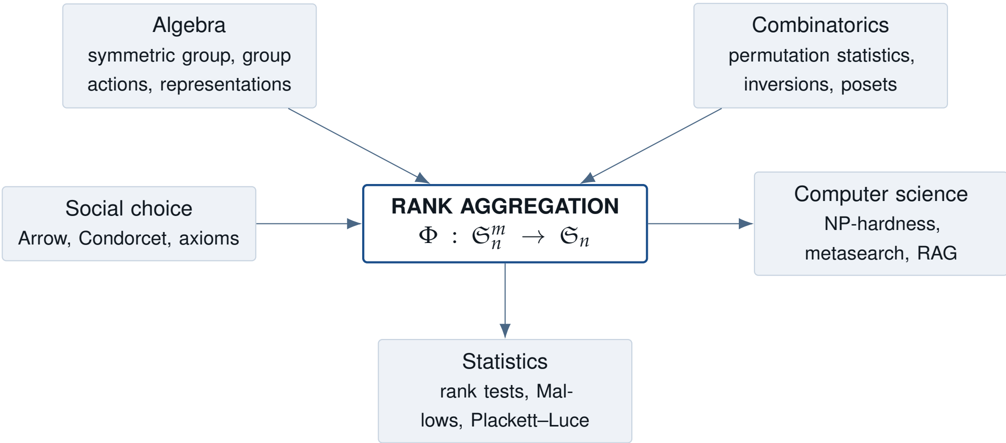
Kenneth Arrow's doctoral thesis converted the subject from a collection of curiosities into a theory with an impossibility theorem at its centre. Arrow's theorem (Chapter 7) says that a short list of obviously desirable properties of an aggregation rule is *inconsistent* for three or more alternatives. Everything after 1951 is, in one sense, the study of which property to give up.

1.2.3 Branch one: combinatorics and algebra

Persi Diaconis, in the 1980s, reframed ranked data as *probability on the symmetric group*, importing representation theory, random walks and metric geometry. This is the branch that gives us Kendall's τ as a word metric, the Diaconis–Graham inequality, and Mallows models. It is the branch closest to the Indian postgraduate syllabus: it is combinatorics and group theory wearing a statistics coat.

1.2.4 Branch two: computer science

From 2001 (Dwork, Kumar, Naor, Sivakumar, *Rank Aggregation Methods for the Web*) the subject became urgent. Metasearch, then recommender systems, then ensemble learning, then — from roughly 2023 — hybrid retrieval inside retrieval-augmented generation systems. This branch contributes complexity theory (Kemeny aggregation is NP-hard) and approximation algorithms, and it contributes Reciprocal Rank Fusion, which is a piece of engineering rather than a piece of mathematics and is none the worse for it.



1.3 Why a mathematician should care

Three reasons, in increasing order of seriousness.

First, it is a place where a small amount of group theory does surprisingly heavy lifting. Theorem 5.12 — that a cyclically symmetric ballot must produce a tie — is a one-line consequence of the orbit structure of a group action, and it simultaneously explains the behaviour of positional rules, pairwise rules and median rules. Few undergraduate applications of group actions are that efficient.

Second, it is a place where *impossibility* is the main theorem. Most of an undergraduate education consists of constructions. Arrow's theorem is a proof that a construction cannot exist, and it has the same flavour as the insolubility of the quintic: a fairness requirement plays the role of a radical extension.

Third, it is a place where the folklore is wrong in an instructive way. The industry belief about RRF is a correct statement about Borda's rule attributed to a formula that in fact does the opposite at the margin. Correcting it requires nothing more than Jensen's inequality, which is exactly the kind of thing a first course in analysis should be for.

PLAIN LANGUAGE

Rank aggregation is the mathematics of *several people handing you ordered lists and asking for one list back*. That is all. The lists might be restaurant preferences, search-engine results, doctors ranking treatments, or three different neural networks ranking documents. The mathematics does not care which; it only cares about the orders.

CHAPTER 2

Rankings as Mathematical Objects

2.1 The symmetric group as a data space

Let $A = \{a_1, \dots, a_n\}$ be a finite set of *alternatives* (candidates, restaurants, documents, treatments). We fix once and for all an identification $A \leftrightarrow [n] := \{1, 2, \dots, n\}$.

Definition 2.1 (Ranking)**.** A *(strict, complete) ranking* of A is a bijection $\sigma : A \to [n]$, where $\sigma(a)$ is the *position* of a. We write $a \succ_\sigma b$ when $\sigma(a) < \sigma(b)$, and read this as "a is preferred to b under σ". The set of all rankings is the symmetric group $\mathfrak{S}_n$, of cardinality $n!$.

> **The one notational trap in the subject.** There are two natural ways to write a permutation, and they are inverse to each other.
>
> - *Rank vector*: $\sigma = (\sigma(a_1), \dots, \sigma(a_n))$ records *where each alternative went*. Small numbers are good.
> - *Order vector* (one-line notation): $\sigma^{-1} = (\sigma^{-1}(1), \dots, \sigma^{-1}(n))$ records *who occupies each position*. This is the ballot as actually written down.
>
> Shailesh's ballot "Vaishali, SP's, Shabree" is an *order vector*. Its rank vector, with alternatives listed alphabetically (Shabree, SP's, Vaishali), is $(3, 2, 1)$. Confusing the two is the commonest error in this subject; several published papers contain it. Throughout this monograph, σ **denotes the rank vector** and $r_\sigma(a) := \sigma(a)$ is the rank of a.

Definition 2.2 (Profile)**.** A *profile* of m rankers is a tuple $P = (\sigma_1, \dots, \sigma_m) \in \mathfrak{S}_n^m$. Two profiles are *anonymous-equivalent* if one is a permutation of the other; a profile considered up to this equivalence is a multiset of rankings.

Definition 2.3 (Aggregation rule)**.** A *social welfare function* (SWF) is a map $\Phi : \mathfrak{S}_n^m \to \mathfrak{S}_n$. A *scoring function* is a map $s : \mathfrak{S}_n^m \to \mathbb{R}^n$; it induces an SWF by sorting s (with some tie-break). We call Φ a *social choice function* if it returns only a winner, $\Phi : \mathfrak{S}_n^m \to A$.

2.2 The two examples we will keep returning to

Example 2.4 (Profile P_1: the original)**.** Alternatives V (Vaishali), S (SP's Biryani), B (Shabree).

$$\sigma_1 : V \succ S \succ B, \qquad \sigma_2 : B \succ S \succ V, \qquad \sigma_3 : S \succ V \succ B.$$

As rank vectors, listing (V, S, B): $\sigma_1 = (1, 2, 3)$, $\sigma_2 = (3, 2, 1)$, $\sigma_3 = (2, 1, 3)$.

Example 2.5 (Profile P_2: the cyclic variant)**.**

$$\tau_1 : V \succ S \succ B, \qquad \tau_2 : S \succ B \succ V, \qquad \tau_3 : B \succ V \succ S.$$

Rank vectors: $\tau_1 = (1, 2, 3)$, $\tau_2 = (3, 1, 2)$, $\tau_3 = (2, 3, 1)$.

Remark 2.6. Stack the rank vectors of P_2 as rows:

$$\begin{pmatrix} 1 & 2 & 3 \\ 3 & 1 & 2 \\ 2 & 3 & 1 \end{pmatrix}.$$

This is a *Latin square*: every symbol occurs once in each row and once in each column. Profiles whose rank-vector array is a Latin square are exactly the profiles on which *nothing can be decided*; this is Theorem 5.12. P_1 is not a Latin square — the column sums are $6, 5, 7$ rather than all equal.

2.3 Weak orders, partial rankings, top-k lists

Real data is rarely a full strict ranking. Three relaxations matter.

Definition 2.7 (Weak order / bucket order)**.** A *weak order* on A is a total preorder: a reflexive, transitive, complete relation $\succeq$ permitting ties. Equivalently, an ordered partition $A = B_1 \sqcup \cdots \sqcup B_t$ into *buckets*, with everything in B_i preferred to everything in B_j for $i < j$.

The number of weak orders on n elements is the *ordered Bell number* (Fubini number) $a_n = \sum_{t \geq 0} t!\, S(n, t)$, where $S(n, t)$ is a Stirling number of the second kind: $1, 1, 3, 13, 75, 541, 4683, \ldots$ Note $a_n \gg n!$; permitting ties enlarges the data space substantially.

Definition 2.8 (Fractional / midrank convention)**.** Given a weak order with buckets $B_1, \ldots, B_t$, the *midrank* of $a \in B_i$ is

$$r(a) \;=\; \Big(\textstyle\sum_{j<i} |B_j| \Big) + \frac{|B_i|+1}{2}.$$

Midranks preserve the total $\sum_a r(a) = \binom{n+1}{2}$, which is what makes rank statistics behave.

Definition 2.9 (Top-k list)**.** A *top-k list* is an injection $\sigma : A_\sigma \to [k]$ defined on a subset $A_\sigma \subseteq A$ with $|A_\sigma| = k$. The elements of $A \setminus A_\sigma$ are *unranked*, which is *not* the same as ranked last.

> The treatment of unranked items is the single most consequential modelling decision in practical rank fusion, and it is almost never discussed. Chapter 5, §5.6 shows that in realistic retrieval settings the handling of unranked documents contributes an order of magnitude more to the final ordering than the choice of the fusion formula itself.

2.4 Group actions: two of them, and do not confuse them

$\mathfrak{S}_n$ acts on rankings in two genuinely different ways, and the distinction is the technical heart of Chapter 5.

Definition 2.10 (Relabelling of alternatives)**.** For $\pi \in \mathfrak{S}_n$ acting on A, define $\pi \cdot \sigma := \sigma \circ \pi^{-1}$. This renames the candidates and leaves the positions alone. A rule Φ is *neutral* if $\Phi(\pi \cdot \sigma_1, \ldots, \pi \cdot \sigma_m) = \pi \cdot \Phi(\sigma_1, \ldots, \sigma_m)$: renaming the candidates renames the answer.

Definition 2.11 (Relabelling of positions)**.** For $\rho \in \mathfrak{S}_n$, define $\sigma \cdot \rho := \rho \circ \sigma$. This shuffles *positions*, which is not usually meaningful — position 1 is better than position 2 and no relabelling should be allowed to deny it. This action is used only as a technical device.

Definition 2.12 (Anonymity)**.** Φ is *anonymous* if $\Phi(\sigma_{\pi(1)}, \ldots, \sigma_{\pi(m)}) = \Phi(\sigma_1, \ldots, \sigma_m)$ for every $\pi \in \mathfrak{S}_m$: the identities of the *voters* do not matter.

> **Neutrality** = candidates are interchangeable. **Anonymity** = voters are interchangeable. Almost every rule used in practice has both. Theorem 5.12 shows that these two harmless-looking symmetries are already enough to force a tie on profile P_2, and hence that no method whatsoever — not RRF, not Borda, not Kemeny, not a neural reranker — can break it without abandoning one of them.

2.5 Counting: how big is the problem?

Proposition 2.13. *The number of anonymous profiles of m rankers over n alternatives is $\binom{n!+m-1}{m}$.*

Proof. An anonymous profile is a multiset of size m drawn from a set of size $n!$; the count is the standard stars-and-bars figure $\binom{n!+m-1}{m}$. ∎

For $n = 3, m = 3$ this is $\binom{8}{3} = 56$; the number of *labelled* profiles is $6^3 = 216$. For $n = 10, m = 3$ the labelled count is $(10!)^3 \approx 4.8 \times 10^{19}$. Exhaustive analysis is available only for tiny n; this is why the subject is axiomatic rather than computational.

Part II

The Geometry of Ranked Data

CHAPTER 3

Metrics on the Symmetric Group

Before we can average rankings we must know what it means for two rankings to be close. There is no single right answer; there are four standard answers, they are mutually comparable, and the comparison theorem (Diaconis–Graham) is one of the prettiest results in the subject.

3.1 Kendall's tau distance

Definition 3.1 (Kendall distance)**.** For $\sigma, \pi \in \mathfrak{S}_n$,

$$d_{\mathrm{K}}(\sigma,\pi) \;:=\; \#\Big\{\{a,b\} \subseteq A \;:\; \big(\sigma(a)-\sigma(b)\big)\big(\pi(a)-\pi(b)\big) < 0\Big\},$$

the number of unordered pairs on which σ and π disagree. Such a pair is a *discordant pair* or an *inversion* of $\sigma\pi^{-1}$.

Proposition 3.2. *d_{K} is a metric on $\mathfrak{S}_n$, with $0 \le d_{\mathrm{K}} \le \binom{n}{2}$, and the upper bound is attained exactly when π is the reversal of σ.*

Proof. Write $d_{\mathrm{K}}(\sigma,\pi) = \sum_{\{a,b\}} \delta_{ab}(\sigma,\pi)$ where $\delta_{ab}(\sigma,\pi) \in \{0,1\}$ records whether the pair is discordant. For a *fixed* pair $\{a,b\}$, the map $\sigma \mapsto \operatorname{sgn}(\sigma(a)-\sigma(b))$ sends $\mathfrak{S}_n$ onto the two-point set $\{+,-\}$, and δ_{ab} is the pullback of the discrete metric on that set. A pullback of a metric along a map is a pseudometric, so each δ_{ab} satisfies the triangle inequality; a finite sum of pseudometrics is a pseudometric. Positivity: if $d_{\mathrm{K}}(\sigma,\pi)=0$ then σ and π agree on every pair, hence induce the same total order, hence are equal. The bound is immediate since there are $\binom{n}{2}$ pairs, with equality iff every pair is discordant, i.e. π reverses σ. ∎

Proposition 3.3 (Word-metric description)**.** *$d_{\mathrm{K}}(\sigma,\pi)$ equals the minimum number of* adjacent transpositions *needed to transform the order vector of σ into that of π. Equivalently, d_{K} is the word metric on the Cayley graph of $\mathfrak{S}_n$ with generating set $\{s_1,\dots,s_{n-1}\}$, $s_i = (i\ i{+}1)$.*

Proof sketch. An adjacent transposition changes the relative order of exactly one pair, so it changes $d_{\mathrm{K}}(\cdot,\pi)$ by exactly ± 1; hence at least $d_{\mathrm{K}}(\sigma,\pi)$ swaps are needed. Conversely, if $\sigma \neq \pi$ there exist two elements adjacent in σ's order whose relative order disagrees with π (otherwise a straightforward induction shows $\sigma=\pi$); swapping them decreases d_{K} by one. Bubble sort realises the bound. ∎

Remark 3.4 (Distribution of d_{K})**.** The generating function for inversions is the *q-factorial*

$$\sum_{\sigma\in\mathfrak{S}_n} q^{d_{\mathrm{K}}(\sigma,\mathrm{id})} \;=\; [n]_q! \;=\; \prod_{i=1}^{n} \frac{1-q^i}{1-q} \;=\; \prod_{i=1}^{n} \big(1+q+\cdots+q^{i-1}\big).$$

Consequently, for σ uniform on $\mathfrak{S}_n$, $d_{\mathrm{K}}(\sigma,\mathrm{id})$ is a sum of independent uniform variables on $\{0,1,\dots,i-1\}$, giving

$$\mathbb{E}[d_{\mathrm{K}}] = \frac{1}{2}\binom{n}{2}, \qquad \operatorname{Var}[d_{\mathrm{K}}] = \frac{n(n-1)(2n+5)}{72},$$

and asymptotic normality. This is the engine behind Kendall's rank-correlation test.

Definition 3.5 (Kendall's τ coefficient)**.** $\tau(\sigma,\pi) = 1 - \dfrac{4\,d_{\mathrm{K}}(\sigma,\pi)}{n(n-1)} \in [-1,1]$, equal to $+1$ for identical rankings and -1 for reversed ones.

3.2 Spearman's footrule and rho

Definition 3.6.

$$d_{\mathrm{F}}(\sigma,\pi) := \sum_{a\in A} \bigl|\sigma(a)-\pi(a)\bigr|, \qquad d_{\mathrm{S}}(\sigma,\pi) := \Bigl(\sum_{a\in A} \bigl(\sigma(a)-\pi(a)\bigr)^2\Bigr)^{1/2}.$$

These are the ℓ^1 and ℓ^2 distances between rank vectors.

Both are obviously metrics (restrictions of ℓ^p metrics on $\mathbb{R}^n$). The diameters are $d_{\mathrm{F}}^{\max} = \lfloor n^2/2 \rfloor$ and $(d_{\mathrm{S}}^{\max})^2 = n(n^2-1)/3$, both attained at the reversal.

Definition 3.7 (Spearman's ρ). $\rho(\sigma,\pi) = 1 - \dfrac{6\sum_a(\sigma(a)-\pi(a))^2}{n(n^2-1)}$, the Pearson correlation of the two rank vectors.

> **Why the footrule matters computationally.** Minimising $\sum_j d_{\mathrm{F}}(\pi,\sigma_j)$ over $\pi \in \mathfrak{S}_n$ is a *bipartite matching* problem — assign alternatives to positions minimising total ℓ^1 cost — and is therefore solvable in $O(n^3)$ by the Hungarian algorithm. Minimising $\sum_j d_{\mathrm{K}}(\pi,\sigma_j)$ is NP-hard (Theorem 8.5). The Diaconis–Graham inequality below converts the easy problem into a provable 2-approximation for the hard one.

3.3 Cayley and Ulam distances

Definition 3.8 (Cayley distance). $d_{\mathrm{C}}(\sigma,\pi)$ is the minimum number of *arbitrary* transpositions carrying σ to π. Equivalently $d_{\mathrm{C}}(\sigma,\pi) = n - c(\sigma\pi^{-1})$, where $c(\cdot)$ is the number of cycles (including fixed points).

Definition 3.9 (Ulam distance). $d_{\mathrm{U}}(\sigma,\pi) = n - \ell(\sigma\pi^{-1})$, where ℓ is the length of the longest increasing subsequence. Equivalently, the minimum number of single-element *move* operations (delete and reinsert anywhere).

Both are right-invariant word metrics, with generating sets "all transpositions" and "all single-element moves" respectively. Their diameters are $n-1$: they are much coarser than d_{K} and d_{F}.

Proposition 3.10 (Right-invariance). *For every $\rho \in \mathfrak{S}_n$ and every $d \in \{d_{\mathrm{K}}, d_{\mathrm{F}}, d_{\mathrm{S}}, d_{\mathrm{C}}, d_{\mathrm{U}}\}$, $d(\sigma\rho,\pi\rho) = d(\sigma,\pi)$.*

This is precisely the statement that these metrics do not care what the alternatives are called — the metric-space avatar of *neutrality*.

3.4 The Diaconis–Graham inequality

Theorem 3.11 (Diaconis–Graham, 1977). *For all $\sigma,\pi \in \mathfrak{S}_n$,*

$$d_{\mathrm{K}}(\sigma,\pi) + d_{\mathrm{C}}(\sigma,\pi) \;\le\; d_{\mathrm{F}}(\sigma,\pi) \;\le\; 2\,d_{\mathrm{K}}(\sigma,\pi).$$

Proof of the right-hand inequality. By right-invariance we may take $\pi = \mathrm{id}$ and write σ for $\sigma\pi^{-1}$. For each a, let $u_a = \#\{b : \sigma(b) < \sigma(a),\ b > a\}$ and $v_a = \#\{b : \sigma(b) > \sigma(a),\ b < a\}$ counting the inversions in which a participates from above and below. Then $\sigma(a) - a = u_a - v_a$, so $|\sigma(a) - a| \le u_a + v_a$. Summing over a and noting that $\sum_a(u_a+v_a) = 2\,d_{\mathrm{K}}(\sigma,\mathrm{id})$ (each inversion is counted at both of its two members) gives $d_{\mathrm{F}} \le 2d_{\mathrm{K}}$. ∎

Metric	Generator	Diameter	Aggregation
Kendall d_{K}	adjacent transposition	$\binom{n}{2}$	NP-hard (Kemeny)
Footrule d_{F}	— (ℓ^1)	$\lfloor n^2/2 \rfloor$	$O(n^3)$ matching
Spearman d_{S}^2	— (ℓ^2)	$n(n^2-1)/3$	$O(n \log n)$ (Borda)
Cayley d_{C}	any transposition	$n-1$	NP-hard
Ulam d_{U}	single-element move	$n-1$	NP-hard

Table 3.1: The five standard metrics on $\mathfrak{S}_n$ and the complexity of finding a median.

The left-hand inequality is harder; see Diaconis and Graham (1977).

Example 3.12 (Exhaustive check, $n = 6$)**.** All 720 permutations were tested against the identity. The inequality holds throughout. The left bound is tight for 511 of the 720 permutations; the right bound is tight for 132.

Corollary 3.13 (Dwork–Kumar–Naor–Sivakumar, 2001)**.** *Let π_{F} minimise $\sum_j d_{\mathrm{F}}(\pi, \sigma_j)$ and let π_{K} minimise $\sum_j d_{\mathrm{K}}(\pi, \sigma_j)$. Then*

$$\sum_j d_{\mathrm{K}}(\pi_{\mathrm{F}}, \sigma_j) \;\le\; 2 \sum_j d_{\mathrm{K}}(\pi_{\mathrm{K}}, \sigma_j).$$

That is, the polynomial-time footrule aggregation is a 2-approximation to the NP-hard Kemeny aggregation.

Proof. Using both halves of Theorem 3.11 and optimality of π_{F} for d_{F}:

$$\sum_j d_{\mathrm{K}}(\pi_{\mathrm{F}}, \sigma_j) \;\le\; \sum_j d_{\mathrm{F}}(\pi_{\mathrm{F}}, \sigma_j) \;\le\; \sum_j d_{\mathrm{F}}(\pi_{\mathrm{K}}, \sigma_j) \;\le\; 2 \sum_j d_{\mathrm{K}}(\pi_{\mathrm{K}}, \sigma_j). \qquad \blacksquare$$

The reason Borda is easy. Minimising $\sum_j d_{\mathrm{S}}(\pi, \sigma_j)^2$ over $\pi \in \mathfrak{S}_n$ is the same as minimising $\sum_a \sum_j (\pi(a) - \sigma_j(a))^2$, which expands to a constant plus $-2 \sum_a \pi(a) \big(\sum_j \sigma_j(a) \big)$. Since $\sum_a \pi(a)^2$ is constant over $\pi \in \mathfrak{S}_n$, the problem reduces to *maximising* $\sum_a \pi(a)\, S(a)$ where $S(a) := \sum_j \sigma_j(a)$. By the rearrangement inequality the maximum occurs when $\pi(a)$ and $S(a)$ are similarly ordered — that is, the best position goes to the alternative with the smallest rank sum. This is exactly *sort by mean rank*, which is Borda's rule. *The ℓ^2 median of a set of rankings is the Borda ranking.*

Example 3.14 (The two profiles, measured)**.** For the cyclic profile P_2 of Example 2.5, every pair of ballots is at the same distance from every other:

$$d_{\mathrm{K}} = 2, \qquad d_{\mathrm{F}} = 4, \qquad d_{\mathrm{C}} = 2 \quad \text{for all three pairs.}$$

The three ballots form an equilateral triangle in $(\mathfrak{S}_n, d_{\mathrm{K}})$. There is no centre to prefer. Note that Diaconis–Graham is tight on both sides here: $2 + 2 = 4 = 2 \cdot 2$.

Part III

Aggregation Rules

CHAPTER 4

Positional Scoring Rules

4.1 Definition and the standard zoo

Definition 4.1 (Positional scoring rule)**.** Fix a *weight vector* $w = (w_1, w_2, \dots, w_n) \in \mathbb{R}^n$ with $w_1 \geq w_2 \geq \cdots \geq w_n$. The induced rule assigns to each alternative

$$s_w(a) \;=\; \sum_{j=1}^{m} w_{\sigma_j(a)},$$

and ranks alternatives by decreasing s_w.

Rule	Weights w_r	Character
Plurality	$\mathbf{1}[r=1]$	only first place counts
Veto (antiplurality)	$\mathbf{1}[r<n]$	only last place counts
k-approval	$\mathbf{1}[r \leq k]$	a threshold
Borda	$n-r$	positions counted linearly
Dowdall / harmonic	$1/r$	strongly top-weighted
Discounted gain (nDCG)	$1/\log_2(r+1)$	the IR evaluation discount
Reciprocal Rank Fusion	$\mathbf{1/(k+r)}$	**our subject**

Table 4.1: Positional rules differ only in the shape of the discount curve.

> Every rule in this table is the *same algorithm* with a different discount curve. Reciprocal Rank Fusion is not a new kind of object; it is a one-parameter family of positional scoring rules. Recognising this immediately gives us the entire classical theory — Young's characterisation, the Condorcet incompatibility, the symmetry obstruction — for free.

4.2 When do two weight vectors give the same rule?

Proposition 4.2 (Affine invariance)**.** *Suppose every ranker ranks every alternative. Then for $\alpha > 0$ and $\beta \in \mathbb{R}$, the weight vectors w and $w' = \alpha w + \beta\mathbf{1}$ induce the same ranking on every profile.*

Proof. $s_{w'}(a) = \sum_j (\alpha w_{\sigma_j(a)} + \beta) = \alpha\, s_w(a) + m\beta$, an increasing affine function of $s_w(a)$ with constants not depending on a. ∎

Corollary 4.3. *Borda's rule ($w_r = n - r$) is identical to "rank by smallest total rank" ($w_r = -r$), to "rank by smallest mean rank", and to the ℓ^2 median of Chapter 3. We shall use these interchangeably.*

> Proposition 4.2 *fails* the moment some ranker omits some alternative. If a document is absent from a list and is assigned score 0 rather than w_n, then adding β to every weight changes the comparison between a ranked and an unranked item. This innocuous-looking failure is, as §5.6 shows, the dominant effect in real retrieval systems.

4.3 Young's characterisation

Definition 4.4 (Reinforcement / consistency)**.** A rule Φ is *consistent* if, whenever two disjoint electorates P and Q both elect a, the merged electorate $P \cup Q$ also elects a.

Theorem 4.5 (Young, 1975)**.** *A social choice correspondence is anonymous, neutral, consistent and continuous if and only if it is a positional scoring rule.*

Consistency is the axiom that pins down the *additive* form $\sum_j w_{\sigma_j(a)}$. Any rule that is not additive across voters — pairwise majority, Kemeny — must violate it, and does.

4.4 Positional rules are never Condorcet-consistent

Definition 4.6 (Condorcet winner)**.** a is a *Condorcet winner* if for every $b \neq a$, a strict majority of rankers place a above b.

Theorem 4.7 (Fishburn)**.** *For $n \geq 3$ alternatives, no positional scoring rule is Condorcet-consistent: for every weight vector w there is a profile possessing a Condorcet winner whom w does not elect.*

Proof for $n = 3$. Write $w = (w_1, w_2, w_3)$; by Proposition 4.2 normalise to $w = (1, t, 0)$ with $t \in [0, 1]$, where $t = (w_2 - w_3)/(w_1 - w_3)$.
Case $t > 0$. Fix an integer $p > 1/t$ and set $q = 2p - 1$. Consider the profile on $\{a, b, c\}$ with $2p + q = 4p - 1$ voters:

$$\begin{array}{ll} p \text{ voters:} & a \succ b \succ c \\ p \text{ voters:} & a \succ c \succ b \\ q \text{ voters:} & b \succ c \succ a \end{array}$$

Pairwise, a beats each of b and c by $2p$ to $q = 2p - 1$, so a is the Condorcet winner. The scores are

$$s(a) = 2p, \qquad s(b) = pt + q = pt + 2p - 1, \qquad s(c) = pt + qt,$$

and $s(b) - s(a) = pt - 1 > 0$ by the choice of p. So the rule elects b, not a.
Case $t = 0$ (plurality). Take 3 voters $a \succ b \succ c$, 2 voters $b \succ c \succ a$, 2 voters $c \succ b \succ a$. Then b beats a by 4 to 3 and c by 5 to 2, so b is the Condorcet winner, while plurality scores are $s(a) = 3 > s(b) = 2 = s(c)$. ∎

Example 4.8 (An explicit failure of RRF_{60})**.** For $k = 60$ the normalised weight is $t = \dfrac{w_2 - w_3}{w_1 - w_3} = \dfrac{1/(62 \cdot 63)}{2/(61 \cdot 63)} = \dfrac{3843}{7812} \approx 0.4919$, so $p = 3$, $q = 5$ suffices. With 11 rankers,

$$3 \times (a \succ b \succ c), \qquad 3 \times (a \succ c \succ b), \qquad 5 \times (b \succ c \succ a),$$

a is the Condorcet winner (6–5 against each rival), but

$$\mathrm{RRF}_{60}(b) = 0.1779734 \ > \ \mathrm{RRF}_{60}(a) = 0.1777257 \ > \ \mathrm{RRF}_{60}(c) = 0.1766513,$$

so RRF_{60} elects b. Borda agrees with RRF here (rank sums $20, 21, 25$), which is Theorem 5.4 again.

Corollary 4.9. *Reciprocal Rank Fusion is not Condorcet-consistent, for any value of k.*

Remark 4.10. Borda's rule does satisfy a weaker guarantee: the Borda winner is never the Condorcet *loser* (an alternative beaten pairwise by all others). Problem B9 asks for a proof, and Problem B10 asks whether RRF inherits this property. It does not, for small k.

CHAPTER 5

Reciprocal Rank Fusion

This is the central chapter. We define RRF, establish its elementary properties, and then prove three theorems that between them replace the folklore account of why it works with an accurate one.

5.1 Definition and provenance

Definition 5.1 (Reciprocal Rank Fusion)**.** Fix $k > 0$. Given rankers $\sigma_1, \ldots, \sigma_m$ (possibly top-k' lists), define

$$\boxed{\mathrm{RRF}_k(a) \;=\; \sum_{j=1}^{m} \frac{1}{k + \sigma_j(a)}}$$

where the sum runs over those j for which a is ranked by σ_j. Alternatives are ordered by decreasing RRF_k.

The rule is due to Cormack, Clarke and Büttcher (SIGIR 2009), who chose $k = 60$ by fitting on a TREC collection and reported that the choice was not delicate. That constant has since been copied into essentially every hybrid-search library in existence, usually without the fitting.

Example 5.2 (Profile P_1, worked)**.** With $k = 60$, weights are $w_1 = 1/61 = 0.0163934$, $w_2 = 1/62 = 0.0161290$, $w_3 = 1/63 = 0.0158730$.

	Shailesh	Rohan	Ashish	RRF_{60}
SP's Biryani (S)	2	2	1	**0.0486515**
Vaishali (V)	1	3	2	0.0483955
Shabree (B)	3	1	3	0.0481394

The winner is S. Note the rank sums: $S \mapsto 5$, $V \mapsto 6$, $B \mapsto 7$. *Borda gives the same order, by a comfortable margin.* Hold that thought.

5.2 Elementary properties

Proposition 5.3. *Fix $k > 0$ and n alternatives, all ranked by all m rankers. Then*

(i) RRF_k *is a positional scoring rule with $w_r = 1/(k + r)$, strictly decreasing and strictly convex in r;*

(ii) $\dfrac{m}{k+n} \le \mathrm{RRF}_k(a) \le \dfrac{m}{k+1}$;

(iii) the total influence of any single ranker is $w_1 - w_n = \dfrac{n-1}{(k+1)(k+n)}$, *which is $O(k^{-2})$;*

(iv) RRF_k *is anonymous, neutral, and monotone (improving a's rank in one list, others fixed, weakly improves a's position in the output).*

Proof. (i) $w'(r) = -(k+r)^{-2} < 0$ and $w''(r) = 2(k+r)^{-3} > 0$. (ii) Immediate. (iii) $\frac{1}{k+1} - \frac{1}{k+n} = \frac{n-1}{(k+1)(k+n)}$. (iv) Anonymity and neutrality hold because the score depends only on the multiset of ranks; monotonicity because w is decreasing. ■

> **What k does, in one line.** The ratio $w_1/w_n = (k+n)/(k+1)$ measures how much louder first place is than last. At $k=0$ this ratio is n; at $k=60$ with $n=3$ it is 1.033; as $k\to\infty$ it tends to 1. *k is a flattening knob.*

k	$w_1 - w_3$	w_1/w_3
0	0.6666667	3.0000
1	0.2500000	2.0000
5	0.0416667	1.3333
10	0.0139860	1.1818
20	0.0041408	1.0952
60	0.0005204	1.0328
100	0.0001923	1.0198
1000	0.0000020	1.0020

Table 5.1: Flattening of the RRF discount curve with k, for $n=3$.

5.3 Theorem I: RRF is a deformation of Borda

Theorem 5.4 (Large-k limit)**.** *Let $S(a) := \sum_{j=1}^{m} \sigma_j(a)$ be the Borda rank-sum and $Q(a) := \sum_{j=1}^{m} \sigma_j(a)^2$. Then as $k\to\infty$,*

$$\mathrm{RRF}_k(a) \;=\; \frac{m}{k} \;-\; \frac{S(a)}{k^2} \;+\; \frac{Q(a)}{k^3} \;+\; O(k^{-4}).$$

Consequently:

(a) *if $S(a)\neq S(b)$, then for all sufficiently large k, RRF_k orders a and b exactly as Borda does;*

(b) *if $S(a)=S(b)$ but $Q(a)\neq Q(b)$, then for all sufficiently large k, RRF_k prefers whichever of a,b has the* larger *Q.*

Proof. Expand $\frac{1}{k+r} = \frac{1}{k}\cdot\frac{1}{1+r/k} = \frac{1}{k}\left(1-\frac{r}{k}+\frac{r^2}{k^2}-\cdots\right)$, valid for $k>r$, and sum over j. Claim (a): the leading term distinguishing a from b is $-(S(a)-S(b))/k^2$, which dominates the $O(k^{-3})$ remainder once k is large. Claim (b): if $S(a)=S(b)$ the k^{-2} terms cancel and the leading difference is $(Q(a)-Q(b))/k^3$. ∎

Remark 5.5 (How large is "sufficiently large"?)**.** Empirically, very small. Simulation over random profiles (restricted to profiles with distinct Borda scores, so that (a) applies) gives the probability that RRF_k reproduces the Borda order exactly:

	$k=0$	$k=1$	$k=5$	$k=10$	$k=20$	$k=60$	$k=200$
$n=5,\ m=3$	0.511	0.565	0.943	1.000	1.000	1.000	1.000
$n=8,\ m=3$	0.089	0.136	0.400	0.661	0.950	1.000	1.000
$n=10,\ m=5$	0.006	0.006	0.079	0.255	0.526	0.994	1.000

> **Consequence, stated bluntly.** On short candidate lists — and after a first-stage retriever has cut to a few dozen documents, the list *is* short — RRF_{60} **is Borda's rule**. The reciprocal is doing essentially no work. Any property you observe of RRF_{60} on a ten-item list is a property of the 1770s.

5.4 Theorem II: the reciprocal favours polarisation, not consensus

We now come to the claim from the diagram in the Preface: that RRF "rewards consistent agreement across rankers over isolated enthusiasm from one". The following theorem says that, holding the Borda score fixed, the reciprocal does the *opposite*.

Theorem 5.6 (Convexity / anti-consensus)**.** *Fix $m, k > 0$ and a rank-sum S. Among all rank vectors $(r_1, \ldots, r_m) \in \{1, \ldots, n\}^m$ with $\sum_j r_j = S$, the RRF score $\sum_j (k + r_j)^{-1}$ is*

(i) minimised *at the most balanced vector (all r_j equal, or as nearly so as integrality permits);*

(ii) maximised *at the most spread vector.*

More generally, if $(r_1, \ldots, r_m)$ majorises $(r'_1, \ldots, r'_m)$ *then* $\sum_j (k + r_j)^{-1} \geq \sum_j (k + r'_j)^{-1}$.

Proof. The map $f(r) = (k + r)^{-1}$ is strictly convex on $(-k, \infty)$. For strictly convex f, the symmetric function $F(r_1, \ldots, r_m) = \sum_j f(r_j)$ is strictly Schur-convex, hence order-preserving with respect to majorisation. The balanced vector is majorised by every other vector with the same sum, and the most spread vector majorises every other; (i) and (ii) follow. The special case "all equal minimises" is Jensen: $\frac{1}{m} \sum_j f(r_j) \geq f(\frac{1}{m} \sum_j r_j)$. ∎

Example 5.7 (Numerical illustration)**.** Two documents, three rankers. Document C ("consistent") is ranked $(2, 2, 2)$; document D ("divisive") is ranked $(1, 1, 4)$. Both have Borda rank-sum 6.

k	$\mathrm{RRF}_k(C)$	$\mathrm{RRF}_k(D)$	winner
1	0.333333333	0.400000000	D
5	0.428571429	0.444444444	D
10	0.250000000	0.253246753	D
60	0.048387097	0.048411885	D
500	0.005976096	0.005976143	D
5000	0.000599760	0.000599760	D

The divisive document wins at every k, by a margin that shrinks like k^{-3} but never changes sign. This is Theorem 5.4(b) and Theorem 5.6 agreeing.

Correcting the folklore. In Example 5.2, SP's Biryani won because its rank-sum was 5 against 6 and 7 — it was the Borda winner outright. Its "consistency" (ranks $2, 2, 1$) was incidental. Had the profile instead given SP's ranks $(1, 1, 4)$ and Vaishali $(2, 2, 2)$, RRF would have selected the polarising SP's, contradicting the stated moral.
The accurate statement is:

> *RRF selects the alternative with the smallest mean rank; where mean ranks tie, it breaks the tie in favour of the more polarising alternative.*

The consensus-seeking behaviour that practitioners correctly observe comes from the rank-sum, i.e. from Borda — and, as the next section shows, from the treatment of unranked items. It does not come from the reciprocal.

5.5 Theorem III: the small-k limit

Proposition 5.8. *As $k \downarrow 0$,* RRF_k *converges to the* Dowdall *(harmonic) rule $w_r = 1/r$. In the further limit of top-weighting, for binary relevance and a single ranker,* RRF_0 *restricted to the best-ranked relevant item is precisely the* reciprocal rank $1/r$ *whose mean over queries is the standard IR metric MRR.*

The one-parameter family, end to end.

$$\underbrace{\text{Dowdall } 1/r}_{k=0} \longrightarrow \underbrace{\text{RRF}_k}_{0<k<\infty} \longrightarrow \underbrace{\text{Borda } -r}_{k\to\infty}$$

Small k: aggressive, top-heavy, a single ranker can impose a winner. Large k: egalitarian, mean-rank-like, no single ranker can. The default $k = 60$ sits, for realistic list lengths, very near the Borda end.

5.6 Theorem IV: where consensus actually comes from

In practice the rankers are truncated: each returns a top-N list, and a document absent from a list contributes 0 rather than w_N. This is the *zero-fill* convention, and it is not a positional scoring rule at all — Proposition 4.2 fails, and the additive structure now encodes *presence* as well as position.

Theorem 5.9 (Zero-fill dominance)**.** *Let m rankers each return top-N lists, and let documents absent from a list score* 0*. Let a be ranked first by one ranker and absent from the other $m-1$; let b be ranked r by all m rankers. Then* $\text{RRF}_k(b) > \text{RRF}_k(a)$ *if and only if*

$$r < m(k+1) - k.$$

Proof. $\text{RRF}_k(a) = 1/(k+1)$ and $\text{RRF}_k(b) = m/(k+r)$. The inequality $m/(k+r) > 1/(k+1)$ rearranges to $m(k+1) > k+r$, i.e. $r < m(k+1) - k$. ■

Example 5.10 (The number that matters)**.** Take the standard configuration $m = 3$ rankers, $k = 60$. The threshold is $r < 3 \cdot 61 - 60 = 123$. That is: **a document sitting at rank 123 in all three lists outranks a document sitting at rank 1 in one list and missing from the other two.**

This is the real consensus mechanism. It is worth roughly a factor of 123 in rank position, whereas the reciprocal's preference for polarisation (Theorem 5.6) is worth $O(k^{-3})$ — a difference of many orders of magnitude. Practitioners who say "RRF rewards documents that several retrievers agree on" are right about the phenomenon and wrong about the cause. The cause is that absent documents score zero, and $m/(k+r)$ grows linearly in the number of lists a document appears in.

Remark 5.11. This also explains the practical sensitivity to *truncation depth* N. Increasing N moves documents from the "absent, score 0" class into the "present, score $\approx 1/(k+N)$" class, a discontinuous jump. Tuning N typically changes results more than tuning k. Very few deployments tune N at all.

5.7 Theorem V: the symmetry obstruction

We can now explain the reader's variant profile completely, and in a way that requires no reference to RRF at all.

Theorem 5.12 (Symmetric profiles are undecidable)**.** *Let $G \leq \mathfrak{S}_n$ act on the alternatives A, and suppose the profile P (as a multiset of ballots) is G-invariant: $g \cdot P = P$ for every $g \in G$. Let Φ be any anonymous, neutral scoring rule $s : \mathfrak{S}_n^m \to \mathbb{R}^A$. Then s is constant on each G-orbit of A. In particular, if G* acts transitively *on A, all alternatives receive identical scores.*

Proof. Let $g \in G$ and $a \in A$. Neutrality gives $s_{g \cdot P}(g(a)) = s_P(a)$; anonymity lets us treat P as a multiset, and G-invariance gives $g \cdot P = P$. Hence $s_P(g(a)) = s_P(a)$. So s_P is constant on the orbit Ga. If the action is transitive there is a single orbit, namely A. ∎

Corollary 5.13 (The reader's profile)**.** *Let P_2 be the profile of Example 2.5 and let $g = (V\ S\ B)$ be the 3-cycle $V \mapsto S \mapsto B \mapsto V$. Then $g \cdot P_2 = P_2$, and $G = \langle g \rangle \cong \mathbb{Z}/3\mathbb{Z}$ acts transitively on $\{V, S, B\}$. Hence every anonymous neutral rule assigns all three restaurants the same score.*

Proof. Apply g to each ballot: $\tau_1 = (V \succ S \succ B) \mapsto (S \succ B \succ V) = \tau_2$, $\tau_2 \mapsto (B \succ V \succ S) = \tau_3$, $\tau_3 \mapsto \tau_1$. So g permutes the ballots and leaves the multiset fixed. Now invoke Theorem 5.12. ∎

> **The important point.** Corollary 5.13 nowhere mentions the formula. It applies to RRF, Borda, plurality, Dowdall, nDCG-weighting, Kemeny, and to any future method a vendor might invent — provided only that the method does not care which voter is which and does not have a favourite restaurant.
> There is no cleverness that resolves P_2. If you want a winner, you must break a symmetry: consult a tie-break, weight the voters unequally, or acquire information the ballots do not contain.

Example 5.14 (Verification)**.** RRF_{60} scores for P_2: each restaurant is ranked once first, once second, once third, so each scores $\frac{1}{61} + \frac{1}{62} + \frac{1}{63} = 0.0483955$. The Borda rank-sums are all 6. The Kemeny costs of the three cyclic orders are all 4 (the three non-cyclic orders cost 5), so Kemeny has three co-optimal answers. Pairwise majority produces $V \succ S \succ B \succ V$, a cycle (Chapter 6).

5.8 What RRF is actually good for

Having removed two false explanations, we owe a true one.

> **The genuine virtue of rank-space fusion.** Suppose ranker 1 is BM25, emitting unbounded positive scores whose scale depends on corpus statistics and query length; ranker 2 is a dense bi-encoder emitting cosine similarities in $[-1, 1]$ tightly clustered near 0.7; ranker 3 is a cross-encoder emitting logits. Naive score addition (*CombSUM*) is meaningless: whichever ranker happens to have the largest dynamic range decides everything. Score normalisation (min–max, z-score) requires per-query calibration and is notoriously unstable when one list is nearly constant.
> Rank-space fusion sidesteps this entirely: *ranks are already commensurable*. RRF requires no normalisation, no calibration, no tuning, and no training data. That is its real contribution, and it is a substantial one.

> **And its real cost.** Discarding magnitude discards *confidence*. A retriever that is certain about its top document is given exactly the same say as one that is guessing between near-identical scores. When one retriever is genuinely much better than the others for a given query, RRF will dilute it. This is the case for learned fusion weights, or for skipping fusion and reranking the union with a cross-encoder.

5.9 Diagnostics: when to distrust your own output

Definition 5.15 (Resolution)**.** Define the *resolution* of RRF_k at depth d as the smallest positive gap between achievable fused scores among documents ranked in the top d by each of m rankers.

For $m = 3$, $k = 60$, $d = 60$, there are 36,791 distinct achievable scores and the smallest positive gap is 1.3×10^{-11}, against a top-of-list gap $w_1 - w_2 = 2.64 \times 10^{-4}$. The score scale is therefore extremely non-uniform.

Practical rule. If the gap between your ith and $(i+1)$th fused scores is smaller than $w_1 - w_2 = \frac{1}{(k+1)(k+2)}$ — that is, smaller than the value of promoting a single document by one position in a single list — then the relative order of those two documents is determined by arithmetic noise, not by evidence. Log this. In a production system it is worth emitting a `fusion_margin` field and alerting when the top-5 margins collapse; that is the signal to escalate to a reranker rather than to retune k.

CHAPTER 6

Pairwise Majority and the Condorcet Cycle

Positional rules aggregate *positions*. The rival tradition aggregates *comparisons*. It has better fairness properties and one catastrophic defect.

6.1 The majority tournament

Definition 6.1 (Majority margin)**.** For a profile P and $a \neq b$, let $N(a,b) := \#\{j : \sigma_j(a) < \sigma_j(b)\}$ and let the *margin* be $\mu(a,b) := N(a,b) - N(b,a)$. The *majority relation* is $a \to b$ iff $\mu(a,b) > 0$.

When m is odd and rankings are strict, $\mu(a,b) \neq 0$ for all pairs and the majority relation is a *tournament*: a complete asymmetric digraph on A.

Example 6.2. For P_1 (Example 2.4): $\mu(S,V) = 1$, $\mu(S,B) = 1$, $\mu(V,B) = 1$. The tournament is transitive: $S \to V \to B$, and S is a Condorcet winner.

For P_2 (Example 2.5): $\mu(V,S) = 1$, $\mu(S,B) = 1$, $\mu(B,V) = 1$. The tournament is the 3-cycle $V \to S \to B \to V$. There is no Condorcet winner.

> **Two different diagnoses of the same ballot.** RRF reports a perfect three-way tie; pairwise majority reports a cycle. Both are correct descriptions of P_2. They are not the same statement: a tie says "the alternatives are equally good", a cycle says "group preference is not an ordering at all". Rank aggregation is a *choice of lens*, not a computation of a pre-existing truth.

6.2 Condorcet's paradox and its ubiquity

Theorem 6.3 (McGarvey, 1953)**.** *Every tournament on n alternatives arises as the majority tournament of some profile of strict rankings. Indeed $n(n-1)$ voters suffice.*

Proof. For each desired edge $a \to b$, add two voters with rankings

$$a \succ b \succ (\text{rest in some order } c_1 \succ \cdots \succ c_{n-2}), \qquad (c_{n-2} \succ \cdots \succ c_1) \succ a \succ b.$$

These two voters contribute $+2$ to $\mu(a,b)$ and cancel on every other pair. Doing this for each of the $\binom{n}{2}$ desired edges yields the tournament. ∎

Corollary 6.4. *No structural assumption on the profile can rule out cycles. The majority relation carries no information beyond "a tournament".*

6.2.1 How likely is a cycle?

Under the *impartial culture* model — each of m voters draws a ranking uniformly and independently from $\mathfrak{S}_n$ — the probability of no Condorcet winner with $n = 3$ can be computed exactly by enumeration:

Voters m	$\mathbb{P}$[no Condorcet winner]	decimal
3	$1/18$	0.055556
5	$5/72$	0.069444
7	$875/11664$	0.075017
∞	$\frac{1}{4} - \frac{3}{2\pi} \arcsin \frac{1}{3}$	0.087740

Theorem 6.5 (Guilbaud's number)**.** *For $n = 3$ alternatives under impartial culture,*

$$\lim_{m \to \infty} \mathbb{P}[\exists \text{ Condorcet winner}] = \frac{3}{4} + \frac{3}{2\pi} \arcsin \frac{1}{3} \approx 0.9123.$$

The proof is a central-limit computation: the three margin variables $(\mu(a, b), \mu(b, c), \mu(c, a))$, suitably scaled, converge to a trivariate normal that is equicorrelated with $\rho = -1/3$, and the limit is an orthant probability evaluated by the arcsine formula $\frac{1}{8} + \frac{3}{4\pi} \arcsin \rho$ (doubled, for the two cycle orientations). The probability of a cycle *increases* with n (to 1 as $n \to \infty$).

> Impartial culture is a worst case, not a realistic one. Real rankers are correlated — BM25 and a dense retriever agree far more often than chance — and correlation suppresses cycles. Under *single-peaked* preferences (Black's theorem), cycles are impossible altogether. The honest summary is: cycles are rare in practice and *cannot be excluded in principle*.

6.3 Living with cycles

Definition 6.6 (Smith set / top cycle)**.** The *Smith set* is the smallest nonempty $S \subseteq A$ such that every member of S beats every non-member pairwise. It always exists and is unique.

When a Condorcet winner exists, the Smith set is the singleton containing it. For P_2 the Smith set is all of $\{V, S, B\}$ — confirming, again, that the ballot contains no information sufficient to narrow the field.

Method	**Rule**	**On P_2**
Copeland	maximise (wins − losses) in the tournament	three-way tie
Minimax	minimise the largest margin of defeat	three-way tie
Ranked Pairs	lock in margins in decreasing order, skipping any that create a cycle	three-way tie
Schulze	maximise the widest "beatpath"	three-way tie
Kemeny	minimise total Kendall distance	three co-optimal orders

Table 6.1: Condorcet-family methods. Every one of them is neutral and anonymous, and so every one of them is bound by Theorem 5.12.

CHAPTER 7

Arrow's Impossibility Theorem

7.1 Statement

Let $L(A)$ denote the set of strict linear orders on A, $|A| = n \geq 3$, and let $F : L(A)^m \to L(A)$ be a social welfare function.

Definition 7.1 (The axioms)**.**

(U) **Unrestricted domain.** F is defined on all of $L(A)^m$.

(P) **Pareto / unanimity.** If every voter ranks $a \succ b$, then so does $F(P)$.

(IIA) **Independence of irrelevant alternatives.** If two profiles P, P' agree on every voter's ranking of the pair $\{a, b\}$, then $F(P)$ and $F(P')$ agree on $\{a, b\}$.

(D) **Dictatorship.** There is a voter i with $F(P) = \sigma_i$ for all P.

Theorem 7.2 (Arrow, 1951)**.** *If $n \geq 3$ and F satisfies (U), (P) and (IIA), then F is a dictatorship.*

7.2 Proof

We give the decisive-coalition proof. Throughout, "profile" means an element of $L(A)^m$ and F satisfies (U), (P), (IIA).

Definition 7.3. Let $S \subseteq N = \{1, \ldots, m\}$ and $a \neq b$.

- S is *almost decisive for* (a, b), written $\bar{D}_S(a, b)$, if: whenever every $i \in S$ ranks $a \succ b$ *and* every $i \notin S$ ranks $b \succ a$, then $F(P)$ ranks $a \succ b$.
- S is *decisive for* (a, b), written $D_S(a, b)$, if: whenever every $i \in S$ ranks $a \succ b$ (with no condition on the others), $F(P)$ ranks $a \succ b$.
- S is *decisive* if $D_S(x, y)$ for all ordered pairs $x \neq y$.

Lemma 7.4 (Field expansion)**.** *If $\bar{D}_S(a, b)$ holds for some pair (a, b), then S is decisive.*

Proof. Let $c \notin \{a, b\}$ (possible since $n \geq 3$).

Step 1: $D_S(a, c)$. Consider any profile in which

- every $i \in S$ ranks $a \succ b \succ c$;
- every $i \notin S$ ranks b first, with a and c below it *in either order*.

By (U) such profiles exist. Since S ranks $a \succ b$ and its complement ranks $b \succ a$, $\bar{D}_S(a, b)$ gives $F(P) : a \succ b$. Every voter ranks $b \succ c$, so (P) gives $F(P) : b \succ c$. Transitivity of $F(P)$ yields $F(P) : a \succ c$. Now the complement's ordering of $\{a, c\}$ was arbitrary, so by (IIA) the conclusion $F(P) : a \succ c$ holds whenever S ranks $a \succ c$, whatever the others do. That is $D_S(a, c)$.

Step 2: $D_S(c, b)$. Consider any profile in which

- every $i \in S$ ranks $c \succ a \succ b$;
- every $i \notin S$ ranks a last, with c and b above it in either order.

Then S ranks $a \succ b$ and the complement ranks $b \succ a$, so $\bar{D}_S(a,b)$ gives $F(P) : a \succ b$; every voter ranks $c \succ a$, so (P) gives $F(P) : c \succ a$; transitivity gives $F(P) : c \succ b$. The complement's ordering of $\{c, b\}$ was arbitrary, so (IIA) yields $D_S(c,b)$.

Step 3: chaining. Steps 1 and 2 upgrade $\bar{D}_S(a,b)$ to $D_S(a,c)$ and $D_S(c,b)$ for every c. Since D implies $\bar{D}$, we may re-apply the steps to the new pairs. Starting from (a,b) we obtain in turn $(a,c), (c,b), (c,a), (b,c), (b,a)$ for each c, and then all pairs by induction on A. Hence S is decisive. ■

Lemma 7.5 (Group contraction)**.** *If S is decisive and $|S| \geq 2$, then some nonempty proper subset of S is decisive.*

Proof. Partition $S = S_1 \sqcup S_2$ with both parts nonempty, and pick three distinct alternatives a, b, c. By (U) choose the profile

$$\begin{array}{ll} i \in S_1: & a \succ b \succ c \\ i \in S_2: & b \succ c \succ a \\ i \notin S: & c \succ a \succ b \end{array}$$

Every member of S ranks $b \succ c$, and S is decisive, so $F(P) : b \succ c$. Now compare a and c in $F(P)$.

Case 1: $F(P) : a \succ c$. In this profile S_1 ranks $a \succ c$ while every voter outside S_1 (that is, S_2 and $N \setminus S$) ranks $c \succ a$. Hence $\bar{D}_{S_1}(a,c)$, and Lemma 7.4 makes S_1 decisive.

Case 2: $F(P) : c \succ a$. With $F(P) : b \succ c$ and transitivity we get $F(P) : b \succ a$. In this profile S_2 ranks $b \succ a$ while everyone outside S_2 ranks $a \succ b$. Hence $\bar{D}_{S_2}(b,a)$, and S_2 is decisive.

In either case a nonempty proper subset of S is decisive. ■

Proof of Theorem 7.2. By (P), the full electorate N is decisive. Apply Lemma 7.5 repeatedly. Since N is finite and each application strictly decreases cardinality, the process terminates at a decisive singleton $\{i\}$. For any $a \neq b$, voter i ranking $a \succ b$ forces $F(P) : a \succ b$; as this holds for every pair, $F(P) = \sigma_i$. So i is a dictator. ■

7.3 Which axiom does each method abandon?

Method	Sacrifices	Symptom
Borda, RRF, all positional rules	IIA	sensitive to adding/removing alternatives
Pairwise majority	transitivity of output	Condorcet cycles
Kemeny	IIA	same as Borda, plus NP-hardness
Single-peaked domains (Black)	unrestricted domain (U)	needs structural assumption
Range/approval voting	ordinal input	requires cardinal information

Example 7.6 (RRF violates IIA, concretely)**.** Take two rankers over $\{a, b\}$ only, $k = 60$: ranker 1 says $a \succ b$, ranker 2 says $b \succ a$. Scores tie. Now insert a third alternative c: ranker 1 says $a \succ b \succ c$, ranker 2 says $b \succ c \succ a$. Now $\mathrm{RRF}(a) = \frac{1}{61} + \frac{1}{63} = 0.0322664$ and $\mathrm{RRF}(b) = \frac{1}{62} + \frac{1}{61} = 0.0325224$, so $b \succ a$. The *relative* order of a and b changed although no voter changed their opinion about a versus b. This is exactly the "irrelevant alternative" effect — and in retrieval it means that **adding a document to the candidate pool can flip the order of two documents already in it**.

> **Engineering consequence.** If your evaluation harness computes fusion over the top-100 union and your production system fuses over the top-50, the two can disagree about documents present in both. This is not a bug in the implementation; it is Arrow's theorem.

> Fix the pool size and record it alongside k.

Remark 7.7 (Two companions)**.** *Gibbard–Satterthwaite (1973–75)*: every non-dictatorial, onto social *choice* function on $n \geq 3$ alternatives is manipulable — some voter can gain by misreporting. *Sen's liberal paradox (1970)*: Pareto is incompatible with granting even two individuals decisive power over one pair each. Both are proved by the same decisive-coalition machinery.

CHAPTER 8

Kemeny–Young: The Median Ranking

8.1 Definition

Definition 8.1 (Kemeny aggregation)**.** Given $P = (\sigma_1, \ldots, \sigma_m)$, a *Kemeny consensus* is any

$$\pi^\star \in \arg\min_{\pi \in \mathfrak{S}_n} \sum_{j=1}^{m} d_K(\pi, \sigma_j).$$

The minimum value is the *Kemeny score* of the profile.

This is the *metric median* of the profile in the metric space $(\mathfrak{S}_n, d_K)$ — the exact analogue of the median of a set of real numbers under $|x - y|$. It is the most principled aggregation rule in the subject and the least practical.

Proposition 8.2 (Margin form)**.** $\sum_{j=1}^{m} d_K(\pi, \sigma_j) = \sum_{(a,b):\, a \succ_\pi b} N(b, a)$*, where* $N(b, a)$ *is the number of rankers preferring* b *to* a*. Hence Kemeny aggregation is the* minimum feedback arc set *problem on the weighted majority tournament, equivalently the* maximum acyclic subgraph *problem.*

Proof. $d_K(\pi, \sigma_j)$ counts pairs on which π and σ_j disagree. Summing over j and exchanging the order of summation, each ordered pair (a, b) with $a \succ_\pi b$ contributes the number of rankers who disagree, namely $N(b, a)$. ∎

Theorem 8.3 (Condorcet consistency)**.** *If a Condorcet winner* $a^\star$ *exists, then* $a^\star$ *is ranked first in every Kemeny consensus. More generally the Kemeny order respects the Smith set.*

Proof of the first claim. Suppose π is optimal but ranks some b above $a^\star$. Let b be the alternative immediately above $a^\star$ in π, and let π' swap them. By Proposition 8.2, the cost changes only through the pair $\{a^\star, b\}$, and the change is $N(b, a^\star) - N(a^\star, b) = -\mu(a^\star, b) < 0$ since $a^\star$ is a Condorcet winner. So π' is strictly better, contradicting optimality. Hence nothing is above $a^\star$. ∎

Theorem 8.4 (Young–Levenglick, 1978)**.** *Kemeny's rule is the unique social welfare function that is neutral, consistent (reinforcing), and Condorcet-consistent.*

Compare Theorem 4.5: consistency plus neutrality plus *continuity* gives positional rules; consistency plus neutrality plus *Condorcet* gives Kemeny. The two families intersect only trivially — which is Theorem 4.7.

8.2 Complexity

Theorem 8.5 (Bartholdi–Tovey–Trick 1989; Dwork–Kumar–Naor–Sivakumar 2001)**.** *Computing a Kemeny consensus is NP-hard. It remains NP-hard for as few as four voters, and the associated decision problem is complete for the class* Θ_2^p*.*

Practical options:

1. **Exact ILP.** Binary variables $x_{ab} \in \{0,1\}$ ($x_{ab} = 1$ meaning $a \succ b$), with antisymmetry $x_{ab} + x_{ba} = 1$ and transitivity $x_{ab} + x_{bc} + x_{ca} \leq 2$ for all triples; minimise $\sum_{a \neq b} N(b,a)\, x_{ab}$. Solvable for n up to a few hundred with a good solver.
2. **Footrule 2-approximation.** Corollary 3.13: solve the $O(n^3)$ bipartite matching instead.
3. **Borda as a heuristic.** Borda is a 5-approximation and, empirically, much better than that.
4. **Local search.** Start from Borda; apply adjacent swaps while they reduce cost. Converges to a locally optimal ranking that is automatically Condorcet-consistent on adjacent pairs.

8.3 The two profiles, resolved

Example 8.6. Kemeny costs for all six orders (writing V, S, B as before):

Order	cost on P_1	cost on P_2
$V \succ S \succ B$	4	**4**
$V \succ B \succ S$	5	5
$S \succ V \succ B$	**3**	5
$S \succ B \succ V$	4	**4**
$B \succ V \succ S$	6	**4**
$B \succ S \succ V$	5	5

On P_1 the Kemeny consensus is unique — $S \succ V \succ B$, agreeing with both RRF and the Condorcet order. On P_2 there are three co-optimal consensuses, namely the three cyclic rotations, exactly as Theorem 5.12 demands: the group $\mathbb{Z}/3\mathbb{Z}$ permutes them.

PLAIN LANGUAGE

Kemeny asks: "what single ordering would require the fewest total complaints, where a complaint is one person disagreeing about one pair?" It is the fairest answer anyone has axiomatised. It is also, provably, too slow to compute exactly at scale — which is why the world uses Borda in a reciprocal costume instead.

Part IV

Probability and Application

CHAPTER 9

Probabilistic Models of Ranked Data

So far the profile has been treated as raw combinatorial data. If instead we posit that the ballots are noisy observations of a latent truth, aggregation becomes *estimation*, and we can ask which rule is a good estimator.

9.1 The Mallows model

Definition 9.1 (Mallows ϕ-model)**.** Fix a central ranking $\sigma_0 \in \mathfrak{S}_n$ and a dispersion $\theta > 0$. Then

$$\mathbb{P}_{\theta,\sigma_0}(\sigma) \;=\; \frac{1}{Z(\theta)}\, e^{-\theta\, d_{\mathrm{K}}(\sigma,\sigma_0)}, \qquad \sigma \in \mathfrak{S}_n.$$

Proposition 9.2 (Normalising constant)**.** $Z(\theta) = \sum_{\sigma\in\mathfrak{S}_n} e^{-\theta d_{\mathrm{K}}(\sigma,\sigma_0)} = \prod_{i=1}^{n} \frac{1-e^{-i\theta}}{1-e^{-\theta}}$*, independent of* σ_0.

Proof. By right-invariance of d_{K}, substitute $\sigma \mapsto \sigma\sigma_0^{-1}$ to reduce to $\sigma_0 = \mathrm{id}$. Then $Z(\theta) = \sum_\sigma q^{d_{\mathrm{K}}(\sigma,\mathrm{id})}$ with $q = e^{-\theta}$, which is the q-factorial $[n]_q! = \prod_{i=1}^{n}(1+q+\cdots+q^{i-1})$ from Chapter 3. ■

Theorem 9.3 (Mallows MLE is Kemeny)**.** *Given i.i.d. observations $\sigma_1,\ldots,\sigma_m$ from a Mallows model with unknown σ_0 and known or unknown $\theta > 0$, the maximum-likelihood estimate of σ_0 is exactly a Kemeny consensus of the profile.*

Proof. The log-likelihood is $\ell(\sigma_0,\theta) = -\theta\sum_{j=1}^{m} d_{\mathrm{K}}(\sigma_j,\sigma_0) - m\log Z(\theta)$. Since Z does not depend on σ_0 and $\theta > 0$, maximising over σ_0 is minimising $\sum_j d_{\mathrm{K}}(\sigma_j,\sigma_0)$. ■

> Theorem 9.3 is the reason the NP-hardness of Kemeny is irritating rather than merely inconvenient. The most natural probabilistic model of noisy rankings has an intractable MLE. Everything else in the subject is, in one way or another, an approximation to this.

9.2 Plackett–Luce and Bradley–Terry

Definition 9.4 (Plackett–Luce)**.** Assign each alternative a positive weight $w_a > 0$. Generate a ranking by sampling without replacement, at each step choosing among the remaining alternatives with probability proportional to their weights. Explicitly, for a ranking with order vector $(a_1,\ldots,a_n)$,

$$\mathbb{P}(a_1 \succ \cdots \succ a_n) \;=\; \prod_{i=1}^{n} \frac{w_{a_i}}{\sum_{l\geq i} w_{a_l}}.$$

Proposition 9.5 (Luce's choice axiom / pairwise marginals)**.** *Under Plackett–Luce, the induced distribution on any subset $B \subseteq A$ is again Plackett–Luce with the same weights. In particular $\mathbb{P}[a \succ b] = \dfrac{w_a}{w_a+w_b}$, the Bradley–Terry model.*

Theorem 9.6 (Borda is a consistent estimator under Plackett–Luce)**.** *Under Plackett–Luce with weights w, the expected rank of a is*

$$\mathbb{E}[\sigma(a)] \;=\; 1 + \sum_{b \neq a} \frac{w_b}{w_a + w_b},$$

which is strictly decreasing in w_a. Hence ordering alternatives by increasing expected rank recovers the true order of the weights, and the Borda rule — the sample version — is a consistent estimator of the Plackett–Luce ranking as $m \to \infty$.

Proof. The rank of a equals 1 plus the number of alternatives placed above it, so $\mathbb{E}[\sigma(a)] = 1 + \sum_{b \neq a} \mathbb{P}[b \succ a]$. Apply Proposition 9.5. Each summand $w_b/(w_a + w_b)$ is strictly decreasing in w_a. Consistency follows from the strong law applied to the sample mean rank. ■

The honest defence of RRF. By Theorem 5.4, RRF_{60} on short lists is Borda; by Theorem 9.6, Borda consistently recovers the true order under the standard probabilistic model of ranked data. *That* is why RRF works. It is a consistent estimator wearing an unusual costume, and it requires no calibration to be one.

Remark 9.7 (MLE for Plackett–Luce)**.** The log-likelihood is concave in $\log w$, and the classical MM / Zermelo iteration

$$w_a \;\leftarrow\; \frac{\#\{\text{wins of } a\}}{\sum_j \sum_{i \leq \sigma_j(a)} \left(\sum_{l : \sigma_j(l) \geq i} w_l \right)^{-1}}$$

converges monotonically. This is the natural home for *weighted* fusion: fit retriever-specific weights rather than guessing them.

9.3 A diagnostic use of the models

Under a Mallows model with a single central ranking, the ranks assigned to a given alternative by different rankers should be *concentrated*. Large dispersion in an item's rank vector is evidence that the rankers do not share a central ranking — that they are measuring different things.

Read this against Theorem 5.6. Precisely when the model says "these rankers disagree about what this item is, so distrust it", the reciprocal weighting gives the item a small bonus. The bonus is tiny ($O(k^{-3})$) and rarely decisive, but it points the wrong way. If you want a rule that penalises dispersion, use $\sum_j f(r_j)$ with f decreasing and *concave*. The simplest choice is $f(r) = -r^2$, i.e. rank items by *increasing* $\sum_j r_j^2$; this is Schur-concave and therefore strictly rewards balanced rank vectors. Note that $-\log(k + r)$ will *not* do: it is decreasing but convex, so it lands on the same side as the reciprocal. Concavity, not the shape of the tail, is the property that matters here (Problem B24).

CHAPTER 10

Application: Hybrid Retrieval

10.1 The setting

A modern retrieval stack answers a query by running several retrievers in parallel:

- **Lexical (BM25).** Scores are unbounded positive reals depending on corpus statistics, document length and query length. Excellent on rare terms, identifiers, names, error codes.

- **Dense bi-encoder.** Cosine similarities in $[-1, 1]$, typically bunched in a narrow band. Excellent on paraphrase and semantic proximity, poor on exact tokens.

- **Learned sparse (SPLADE and relatives).** Expanded-term dot products; another unbounded scale.

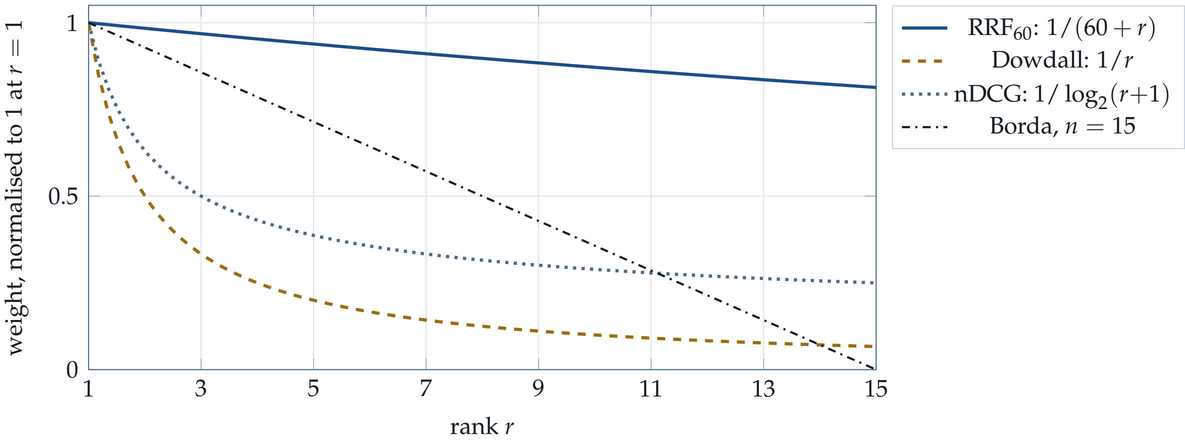
Each returns a top-N list. The problem is to combine them.

10.2 Why score fusion fails and rank fusion does not

CombSUM. Add raw scores. Whichever retriever has the largest dynamic range wins every query. Useless without normalisation.
Normalised CombSUM. Min–max or z-score per query. Better, but unstable: when one list is nearly constant (common for dense retrievers on a homogeneous corpus), the denominator collapses and noise is amplified to full scale. Requires per-query computation and behaves badly on short lists.
RRF. Uses only positions. Nothing to normalise, nothing to calibrate, no training data, one hyperparameter that provably does very little (Table 5.2). This robustness is the whole product.

10.3 Reading the discount curves side by side

The evaluation metric nDCG uses discount $1/\log_2(r+1)$; RRF uses $1/(k+r)$; Borda uses $n-r$. All three are decreasing; the first two are convex.



r	$1/(60+r)$ (norm.)	$1/r$ (norm.)	$1/\log_2(r+1)$ (norm.)	Borda $n-r$, $n=15$ (norm.)
1	1.000	1.000	1.000	1.000
2	0.984	0.500	0.631	0.929
3	0.968	0.333	0.500	0.857
5	0.938	0.200	0.387	0.714
10	0.871	0.100	0.289	0.357
15	0.813	0.067	0.250	0.000

> Look at the second column. Over the entire top-15, the RRF weight falls by less than 19%. Over the same range the Dowdall weight falls by 93% and the nDCG discount by 75%. RRF$_{60}$ **is very nearly a flat weighting**: on a short list it is close to counting how many retrievers returned the document at all, with a mild linear correction. This is Theorem 5.9 and Theorem 5.4 restated as a picture.

10.4 Choosing the parameters

1. **Truncation depth** N**.** The most consequential parameter and the least tuned. Larger N moves documents from "absent, score 0" to "present, small score", a discontinuous change (§5.6). Use the same N for every retriever, and use the same N in evaluation as in production.
2. k**.** Increasing k flattens toward Borda and toward pure appearance-counting; decreasing k toward 0 makes a single retriever's top hit decisive. If your retrievers differ greatly in quality, small k plus explicit weights is more honest than large k.
3. **Number of retrievers** m**.** Theorem 5.9's threshold $r < m(k+1) - k$ grows linearly in m: each additional retriever makes "appearing everywhere" more dominant. Adding a weak fourth retriever can degrade results by diluting the strong ones — this is the RRF version of Arrow's IIA failure.
4. **Weighted RRF.** $\sum_j \lambda_j/(k+\sigma_j(a))$ deliberately abandons anonymity. This is a legitimate and often correct move: your retrievers really are not equally trustworthy. Fit λ by Plackett–Luce MLE or by grid search on a labelled set.

10.5 Evaluation

Definition 10.1. For a query with graded relevance rel_i at position i:

$$\mathrm{DCG@}K = \sum_{i=1}^{K} \frac{2^{\mathrm{rel}_i} - 1}{\log_2(i+1)}, \qquad \mathrm{nDCG@}K = \frac{\mathrm{DCG@}K}{\mathrm{IDCG@}K},$$

MRR $= \mathbb{E}[1/\text{rank of first relevant}]$, and Recall@$K$ is the fraction of relevant documents in the top K.

> Fusion is usually a *recall* intervention, not a precision one. Its job is to get the right document into the candidate pool; a cross-encoder's job is to put it first. Evaluating a fusion change with nDCG@5 while a reranker sits downstream will show noise. Evaluate fusion with Recall@K at the reranker's input width.

10.6 A diagnostic protocol

Given a fused list, three checks are worth running in production. All are cheap.

1. **Margin check.** Compute consecutive score gaps in the top K. Flag any gap below $\frac{1}{(k+1)(k+2)}$, the value of one position in one list. Below that threshold the ordering is arithmetic noise. In the standard configuration $k = 60$ this threshold is 2.64×10^{-4}.
2. **Cycle check.** Build the majority tournament on the top K documents. If the Smith set has size > 1, the retrievers genuinely disagree in a way no fusion rule can settle; escalate.
3. **Symmetry check.** If the rank matrix restricted to the top K is close to a Latin square, Theorem 5.12 applies approximately and the fused ordering is close to arbitrary. This is exactly the reader's restaurant profile appearing in a search index.

> **What to do when a check fires.** Do not retune k; Table 5.2 shows there is almost nothing there to tune. Do one of:
>
> - *Rerank.* Send the union to a cross-encoder, which can read query and document jointly and emit a genuine magnitude rather than a position.
> - *Add information.* Recency, authority, click signal, structural filters — anything the ballots do not contain.
> - *Break the symmetry deliberately and record that you did.* A documented tie-break is respectable; an undocumented one is a reproducibility bug.

10.7 Summary of the practical claims

Common belief	**What the mathematics says**
RRF rewards consensus across rankers over enthusiasm from one.	Partly right, wrong mechanism. The consensus force is zero-fill for absent documents (Thm. 5.9); at equal rank-sum the reciprocal favours the *polarising* item (Thm. 5.6).
$k = 60$ is a carefully chosen constant.	It is a flattening knob fitted once on a 2009 TREC collection. At $k = 60$ on short lists RRF is indistinguishable from Borda (Thm. 5.4).
RRF needs no normalisation, which is why it is robust.	Correct, and this is its genuine and substantial advantage.
More retrievers is better.	Each added retriever increases the dominance of mere appearance (Thm. 5.9) and can flip pairs already present (Arrow / IIA).
A tie in the fused scores means the documents are equally good.	It may instead mean the profile is symmetric and *no* rule can decide (Thm. 5.12).

Part V

Problems

CHAPTER 11

Problem Sets

Set A is computational and should be done with a pen and no software. Set B is proof-based and forms the examinable core. Set C is open-ended; several problems there are small research projects.

Throughout, the *working profile* $P_\star$ over $A = \{a, b, c, d\}$ with three rankers is

$$\sigma_1 : d \succ a \succ c \succ b, \qquad \sigma_2 : a \succ d \succ b \succ c, \qquad \sigma_3 : d \succ b \succ a \succ c.$$

Set A — Foundations

A1. Let σ have order vector $(3, 1, 4, 5, 2)$ on $\{1, \ldots, 5\}$ and let $\pi = \mathrm{id}$. Compute d_{K}, d_{F}, d_{C} and d_{U} between them. Write down the rank vector of σ first.

A2. Verify the Diaconis–Graham inequality for the pair in A1. Which side is tight?

A3. Compute $\mathrm{RRF}_{60}(x)$ for each $x \in \{a, b, c, d\}$ on $P_\star$, to seven decimal places, and give the fused order.

A4. Compute the Borda rank-sums on $P_\star$. Does Borda agree with A3?

A5. Build the majority tournament of $P_\star$. Is there a Condorcet winner? A Condorcet loser?

A6. Compute the Kemeny cost of the order $d \succ a \succ b \succ c$ on $P_\star$, and verify using Proposition 8.2 that this equals $\sum_{x \succ_\pi y} N(y, x)$.

A7. Show that the number of weak orders on a 3-element set is 13, by listing them by bucket type.

A8. For RRF_k with n alternatives, the ratio w_1/w_n equals $(k + n)/(k + 1)$. Evaluate it for $(k, n) = (60, 10)$, $(60, 1000)$, $(0, 10)$ and $(10, 10)$. Comment.

A9. Using Theorem 5.9 with $k = 60$, find the threshold rank $m(k + 1) - k$ for $m = 2, 3, 4, 5$.

A10. Write the rank-vector array of the profile P_2 of Example 2.5. Verify it is a Latin square, and state what Theorem 5.12 then guarantees.

A11. On $P_\star$, compute $d_{\mathrm{K}}(\sigma_1, \sigma_2)$ and convert it to Kendall's τ coefficient.

A12. For rank vectors $u = (1, 2, 3, 4, 5)$ and $v = (2, 1, 4, 5, 3)$, compute Spearman's ρ.

A13. Write the weight vector of 3-approval on $n = 6$ alternatives. Give a positive-affine transformation of it that has $w_6 = 0$ and $w_1 = 1$.

A14. Construct a profile on three alternatives where plurality and Borda elect different winners. Use as few voters as you can.

A15. How many anonymous profiles are there with $n = 3$ alternatives and $m = 4$ rankers?

A16. Expand the q-factorial $[3]_q!$ and interpret each coefficient combinatorially.

A17. Compute $\mathbb{E}[d_{\mathrm{K}}(\sigma, \mathrm{id})]$ and $\mathrm{Var}[d_{\mathrm{K}}(\sigma, \mathrm{id})]$ for σ uniform on $\mathfrak{S}_5$.

A18. With $m = 3$ rankers and $k = 60$: what is the largest rank r such that a document at rank r in *all three* lists outranks a document at rank 1 in one list and absent from the other two?

A19. Six items are placed in buckets $B_1 = \{p\}$, $B_2 = \{q,r,s\}$, $B_3 = \{t,u\}$. Compute all six midranks and verify $\sum r = 21$.

A20. Compute d_U between $(1,2,3,4,5)$ and $(5,4,3,2,1)$, and explain why the Ulam diameter of $\mathfrak{S}_n$ is $n-1$ rather than $\binom{n}{2}$.

Set B — Core

B1. Prove that d_K, d_F, d_C and d_U are all right-invariant, and explain the connection to neutrality.

B2. Prove $d_F \leq 2d_K$ without consulting Chapter 3, and characterise the permutations for which equality holds when $\pi = \mathrm{id}$.

B3. Prove that the diameter of $(\mathfrak{S}_n, d_F)$ is $\lfloor n^2/2 \rfloor$.

B4. Prove that $\sum_{a\in A} \sigma(a) = \binom{n+1}{2}$ for every $\sigma \in \mathfrak{S}_n$, and deduce that no positional rule can be manipulated by a voter who reverses their entire ballot in a way that changes the sum of all scores.

B5. Prove Proposition 4.2. Then give an explicit two-ranker, three-document example with zero-fill in which w and $w + \mathbf{1}$ produce different orders.

B6. Prove that RRF_k is monotone: if a improves by one position in one list and nothing else changes, a's position in the fused order does not worsen.

B7. Prove Theorem 5.6 for $m = 2$ directly, by showing that $g(t) := \frac{1}{k+S/2+t} + \frac{1}{k+S/2-t}$ is strictly increasing in $t \geq 0$.

B8. Prove part (b) of Theorem 5.4, and give an explicit pair of rank vectors and a value of k for which the second-order term decides the order.

B9. Prove that the Borda winner is never the Condorcet loser. *Hint:* show $s_B(a) = \sum_{b\neq a} N(a,b)$ and compare with the average score.

B10. Show that a Latin-square profile has all Borda scores equal, and deduce that it has no Condorcet winner. Verify by computer that for $n = m = 3$ the converse also holds (there are exactly 12 such profiles out of 216), and find a counterexample to the converse for $n = m = 4$.

B11. Prove McGarvey's theorem (Theorem 6.3) and count the voters your construction uses.

B12. Prove that Kemeny's rule respects the Smith set: every member of the Smith set is ranked above every non-member in any Kemeny consensus.

B13. Prove the margin form of the Kemeny objective (Proposition 8.2).

B14. Prove that the Mallows normalising constant $Z(\theta)$ does not depend on σ_0, and evaluate $Z(\theta)$ for $n = 3$.

B15. Prove Theorem 9.3. Where is $\theta > 0$ used?

B16. Prove Proposition 9.5: the pairwise marginal of Plackett–Luce is Bradley–Terry.

B17. Prove Theorem 9.6 and compute $\mathbb{E}[\sigma(a)]$ explicitly for $n = 3$, $w = (3,2,1)$.

B18. Prove Theorem 5.12 in the special case of positional scoring rules, without using the general neutrality argument — i.e. by a direct count of how often each alternative occupies each position.

B19. List all subgroups of $\mathfrak{S}_3$ and determine which act transitively on $\{1,2,3\}$. Which of these can arise as the symmetry group of a profile in the sense of Theorem 5.12?

B20. Construct a profile in which adding a fourth alternative reverses the RRF_{60} order of two of the original three, and identify which Arrow axiom this violates.

B21. Prove Corollary 3.13 in full, stating exactly where each half of Diaconis–Graham is used.

B22. Prove that minimising $\sum_j d_S(\pi, \sigma_j)^2$ over $\pi \in \mathfrak{S}_n$ yields the Borda ranking. *Hint:* rearrangement inequality.

B23. In the Kemeny ILP, show that the constraints $x_{ab} + x_{ba} = 1$ together with $x_{ab} + x_{bc} + x_{ca} \leq 2$ over all ordered triples force x to encode a total order.

B24. Show that $f(r) = -\log(k + r)$ is decreasing but *convex*, so $\sum_j f(r_j)$ is Schur-convex and shares RRF's preference for dispersion. Then show that $f(r) = -r^2$ is decreasing and concave, giving a Schur-concave rule; compare it with RRF_{60} on the rank vectors $(2, 2, 2)$ and $(1, 1, 4)$. What practical drawback does the squared-rank rule have on long lists?

Set C — Advanced and Open-Ended

C1. For which triples (m, n, k) does RRF_k induce exactly the Borda order on *every* profile with distinct Borda scores? Give a sufficient condition on k in terms of m and n, and compare it with the simulation table in Chapter 5.

C2. Derive Guilbaud's number. Set up the trivariate CLT for the margin vector under impartial culture, identify the correlation structure, and evaluate the orthant probability.

C3. Estimate by simulation the probability of a Condorcet cycle under impartial culture for $n = 3, \ldots, 10$ with $m = 51$. Fit the growth and compare with the known result that the probability tends to 1.

C4. Design a fusion rule that is (i) rank-based, hence scale-free, (ii) Schur-concave, hence consensus-preferring, and (iii) reduces to something sensible under zero-fill. Evaluate it against RRF_{60} on a public retrieval benchmark.

C5. Fit weighted RRF, $\sum_j \lambda_j/(k + r_j)$, by treating λ as Plackett–Luce weights over rankers. Derive the MM update and discuss identifiability.

C6. Borda is known to be a 5-approximation to Kemeny. Is RRF_k? Prove a bound, or exhibit a family of profiles on which the ratio grows.

C7. Define the Fagin–Kumar–Sivakumar generalisation of d_K to top-k lists. Show it is a near-metric (triangle inequality up to a constant) and relate its parameter to the zero-fill convention of §5.6.

C8. The Ulam distance from the identity is n minus the longest increasing subsequence. State the Baik–Deift–Johansson theorem on the fluctuations of the LIS and explain what it says about the geometry of $(\mathfrak{S}_n, d_U)$ for large n.

C9. The space $\mathbb{R}^{\mathfrak{S}_n}$ carries a natural $\mathfrak{S}_n$-action. Decompose it into irreducibles and show that the Borda score is the projection onto the $(n-1, 1)$ component. What does the reciprocal weighting add?

C10. Design an experiment on a live retrieval system that isolates the zero-fill effect (Theorem 5.9) from the reciprocal weighting. What do you vary, what do you hold fixed, and what would falsify the claim that zero-fill dominates?

C11. Show that Arrow's theorem fails for $n = 2$ by exhibiting a non-dictatorial rule satisfying (U), (P), (IIA). Where exactly does the proof of Lemma 7.4 break?

C12. Read Kirman–Sondermann (1972) and explain how Arrow's theorem becomes, for infinite electorates, the statement that decisive coalitions form an ultrafilter — and why a dictator is precisely a principal ultrafilter.

C13. State and prove Black's median-voter theorem for single-peaked preferences. Then decide: is there a natural single-peakedness condition on retrieval scores, and would it be testable?

C14. Implement exact Kemeny by ILP for $n \leq 40$ and compare its output with RRF_{60} on a hybrid-retrieval candidate pool. Report the Kendall distance between the two outputs as a function of pool size.

Marginal notes for examination candidates

> **CSIR-NET / GATE MA relevance.** The directly examinable material is: cycle structure and conjugacy in $\mathfrak{S}_n$ (A1, B1, B19); counting permutations by inversions and the q-factorial (A16, A17); metric-space axioms and verification (B1–B3); Jensen and Schur convexity (B7, B24); the rearrangement inequality (B22); group actions, orbits and transitivity (B18, B19); ordered Bell numbers and Stirling numbers (A7); and elementary MLE (B15, B16). Problems A5–A6, B9–B13 are excellent training in careful combinatorial argument even though voting theory itself is not on the syllabus.

CHAPTER 12

Solutions

Set A

A1. The order vector is $(3,1,4,5,2)$, so element 1 sits in position 2, element 2 in position 5, element 3 in position 1, element 4 in position 3, element 5 in position 4: the rank vector is $(2,5,1,3,4)$. *Kendall:* the inversions of the order vector are the pairs $(3,1),(3,2),(4,2),(5,2)$, so $d_K = 4$. *Footrule:* $|2-1|+|5-2|+|1-3|+|3-4|+|4-5| = 1+3+2+1+1 = 8$. *Cayley:* the permutation $1 \mapsto 2 \mapsto 5 \mapsto 4 \mapsto 3 \mapsto 1$ is a single 5-cycle, so $c = 1$ and $d_C = 5-1 = 4$. *Ulam:* the longest increasing subsequence of $(3,1,4,5,2)$ is $3,4,5$ of length 3, so $d_U = 5-3 = 2$.

A2. $d_K + d_C = 4+4 = 8 = d_F$ and $2d_K = 8 = d_F$. *Both* sides are tight simultaneously — a coincidence permitted by the inequality but not typical.

A3. Ranks: $d \mapsto (1,2,1)$, $a \mapsto (2,1,3)$, $b \mapsto (4,3,2)$, $c \mapsto (3,4,4)$. Hence

$$\mathrm{RRF}_{60}(d) = 0.0489159, \quad \mathrm{RRF}_{60}(a) = 0.0483955, \quad \mathrm{RRF}_{60}(b) = 0.0476270, \quad \mathrm{RRF}_{60}(c) = 0.0471230,$$

giving $d \succ a \succ b \succ c$.

A4. Rank sums: $d = 4$, $a = 6$, $b = 9$, $c = 11$. Borda gives $d \succ a \succ b \succ c$, identical to A3 — as Theorem 5.4 predicts for a four-item list at $k = 60$.

A5. Pairwise wins out of three: d beats a (2–1), b (3–0) and c (3–0); a beats b (2–1) and c (3–0); b beats c (2–1). The tournament is transitive: $d \to a \to b \to c$. The Condorcet winner is d; the Condorcet loser is c.

A6. Direct count gives cost 3. By the margin form, summing $N(y,x)$ over ordered pairs $x \succ_\pi y$ in $d \succ a \succ b \succ c$: $N(a,d) = 1$, $N(b,d) = 0$, $N(c,d) = 0$, $N(b,a) = 1$, $N(c,a) = 0$, $N(c,b) = 1$; total 3. This is the minimum, so $d \succ a \succ b \succ c$ is the Kemeny consensus and all four methods agree on $P_\star$.

A7. By number of buckets: one bucket, $1!\,S(3,1) = 1$; two buckets, $2!\,S(3,2) = 2 \cdot 3 = 6$; three buckets, $3!\,S(3,3) = 6$. Total $1+6+6 = 13$, the third ordered Bell number.

A8. $(60,10): 70/61 = 1.1475$. $(60,1000): 1060/61 = 17.377$. $(0,10): 10$. $(10,10): 20/11 = 1.818$. Comment: at $k = 60$ on a short list the discount curve is nearly flat, so RRF is close to counting appearances; the large ratio at $n = 1000$ is generated entirely in the deep tail, where the weights are individually negligible.

A9. $m(k+1) - k = 61m - 60$: for $m = 2,3,4,5$ this is 62, 123, 184, 245.

A10. With alternatives ordered (V,S,B) the rank-vector rows are $(1,2,3)$, $(3,1,2)$, $(2,3,1)$. Each column contains $\{1,2,3\}$, so this is a Latin square. It is moreover invariant under the 3-cycle $(V\,S\,B)$, which acts transitively, so Theorem 5.12 forces every anonymous neutral rule to assign all three alternatives the same score.

A11. The pairs $\{a,d\}$ and $\{b,c\}$ are discordant between σ_1 and σ_2; the other four agree. So $d_K = 2$ and $\tau = 1 - \frac{4 \cdot 2}{4 \cdot 3} = \frac{1}{3}$.

A12. Differences $-1,1,-1,-1,2$, squares summing to 8. Then $\rho = 1 - \frac{6 \cdot 8}{5 \cdot 24} = 1 - 0.4 = 0.6$.

A13. $w = (1, 1, 1, 0, 0, 0)$. It already satisfies $w_1 = 1$, $w_6 = 0$, so it is its own normalised representative; any $\alpha w + \beta \mathbf{1}$ with $\alpha > 0$ induces the same rule. Note w is only *weakly* decreasing — positional rules do not require strict decrease.

A14. Four voters suffice: $2 \times (a \succ b \succ c)$, $1 \times (b \succ c \succ a)$, $1 \times (c \succ b \succ a)$. Plurality: $a = 2$, $b = 1$, $c = 1$, so a wins. Borda with $w = (2, 1, 0)$: $a = 4$, $b = 5$, $c = 3$, so b wins. Three voters cannot suffice: a strict plurality winner must hold at least two first places, giving Borda score ≥ 4, while any rival holds at most one first and two seconds, giving at most 4. So no rival can strictly beat the plurality winner on Borda.

A15. $\binom{3!+4-1}{4} = \binom{9}{4} = 126$.

A16. $[3]_q! = 1 \cdot (1+q)(1+q+q^2) = 1 + 2q + 2q^2 + q^3$. The coefficient of q^t is the number of permutations of $\{1, 2, 3\}$ with exactly t inversions: $1, 2, 2, 1$ for $t = 0, 1, 2, 3$, summing to $3! = 6$.

A17. $\mathbb{E}[d_{\mathrm{K}}] = \frac{1}{2}\binom{5}{2} = 5$ and $\mathrm{Var} = \frac{5 \cdot 4 \cdot 15}{72} = \frac{300}{72} = \frac{25}{6} \approx 4.1667$.

A18. We need $\frac{3}{60+r} > \frac{1}{61}$, i.e. $r < 123$. The largest integer is $r = 122$.

A19. $r(p) = 0 + \frac{1+1}{2} = 1$; $r(q) = r(r) = r(s) = 1 + \frac{3+1}{2} = 3$; $r(t) = r(u) = 4 + \frac{2+1}{2} = 5.5$. Sum $= 1 + 9 + 11 = 21 = \binom{7}{2}$.

A20. The longest increasing subsequence of $(5, 4, 3, 2, 1)$ has length 1, so $d_{\mathrm{U}} = 4 = n - 1$. A single *move* can place any element into its final position, so $n - 1$ moves always suffice (the last element is then correct automatically); hence the diameter is at most $n - 1$, and the reversal attains it. Kendall's generator (adjacent transposition) is far weaker — it changes exactly one pair — which is why its diameter is quadratic.

Set B

B1. For d_{K}: the pair $\{a, b\}$ is discordant for $(\sigma\rho, \pi\rho)$ exactly when $\{\rho(a), \rho(b)\}$ is discordant for (σ, π), and ρ induces a bijection on unordered pairs, so the counts agree. For d_{F} and d_{S}, reindex the sum by $a' = \rho(a)$. For d_{C} and d_{U}, note $(\sigma\rho)(\pi\rho)^{-1} = \sigma\pi^{-1}$, so the defining cycle and subsequence statistics are unchanged. Connection: right-invariance is the metric-level form of neutrality. Any rule defined as a metric median with a right-invariant metric — Kemeny, footrule aggregation — is automatically neutral.

B2. As in Chapter 3: with $\pi = \mathrm{id}$, write $u_a = \#\{b > a : \sigma(b) < \sigma(a)\}$ and $v_a = \#\{b < a : \sigma(b) > \sigma(a)\}$. Then $\sigma(a) - a = u_a - v_a$, so $|\sigma(a) - a| \leq u_a + v_a$, and $\sum_a (u_a + v_a) = 2d_{\mathrm{K}}$. Equality requires $|u_a - v_a| = u_a + v_a$ for every a, i.e. $u_a v_a = 0$: *no element participates in inversions from both sides.*

B3. Take $\pi = \mathrm{id}$. Since $|x - y| = 2\max(x, y) - x - y$ and $\sum_a \sigma(a) = \sum_a a$, we get $d_{\mathrm{F}} = 2\sum_a \max(\sigma(a), a) - n(n+1)$. The values $\max(\sigma(a), a)$ are n entries of the multiset $\{1, 1, 2, 2, \ldots, n, n\}$, so their sum is at most the sum of the n largest entries of that multiset. For n even that bound is $n(n+1) - \frac{n(n+2)}{4}$, giving $d_{\mathrm{F}} \leq n^2/2$; the permutation exchanging $\{1, \ldots, n/2\}$ with $\{n/2+1, \ldots, n\}$ attains it. The odd case is identical with a floor.

B4. σ is a bijection $A \to [n]$, so $\sum_a \sigma(a) = \sum_{r=1}^{n} r = \binom{n+1}{2}$. Consequently the *total* positional score $\sum_a s_w(a) = \sum_j \sum_{r=1}^{n} w_r = m \sum_r w_r$ is independent of the profile: a positional rule is a *conservation law*. A voter can redistribute score among alternatives but cannot create or destroy any.

B5. The first part is Proposition 4.2. For the failure under zero-fill, take two rankers each returning a top-2 list over $\{x, y, z\}$:

$$R_1 : y \succ x, \qquad R_2 : z \succ x,$$

with the missing alternative scoring 0. With $w = (1, 0)$: $s(x) = 0$, $s(y) = s(z) = 1$, so $y, z \succ x$. With $w' = w + 2 \cdot \mathbf{1} = (3, 2)$: $s(x) = 2 + 2 = 4$, $s(y) = s(z) = 3$, so $x \succ y, z$. A positive-affine shift has reversed the order. The reason: x is *present* in both lists while y and z are each present in only one, and raising the baseline rewards presence.

B6. Suppose a moves from rank r to $r - 1$ in list j, exchanging with b. Then $\mathrm{RRF}_k(a)$ increases by $\frac{1}{k+r-1} - \frac{1}{k+r} > 0$, $\mathrm{RRF}_k(b)$ decreases by the same amount, and all other scores are unchanged. Since a's score strictly increases and no other score increases, a's position in the sorted output cannot worsen.

B7. Write $c = k + S/2$ and let the two ranks be $c - k \pm t$. Then

$$g(t) = \frac{1}{c+t} + \frac{1}{c-t} = \frac{2c}{c^2 - t^2},$$

which is strictly increasing on $0 \le t < c$ because the denominator strictly decreases. So among rank pairs with a fixed sum, the more spread pair scores strictly higher.

B8. If $S(a) = S(b)$ the k^{-2} terms in Theorem 5.4 cancel and

$$\mathrm{RRF}_k(a) - \mathrm{RRF}_k(b) = \frac{Q(a) - Q(b)}{k^3} + O(k^{-4}),$$

so the sign is that of $Q(a) - Q(b)$ for large k. Example: $(1, 1, 4)$ and $(2, 2, 2)$ both have $S = 6$, with $Q = 18$ and 12. At $k = 60$ the predicted gap is $6/60^3 = 2.78 \times 10^{-5}$; the exact gap is $0.048411885 - 0.048387097 = 2.479 \times 10^{-5}$.

B9. With $w_r = n - r$, for each voter j we have $n - \sigma_j(a) = \#\{b : a \succ_j b\}$, so summing over j gives $s_B(a) = \sum_{b \neq a} N(a, b)$. Summing over a, $\sum_a s_B(a) = m\binom{n}{2}$, so the average score is $m(n-1)/2$. If a is the Condorcet loser then $N(a, b) < m/2$ for every $b \neq a$, whence $s_B(a) < (n-1)m/2$ — strictly below average. Some alternative must be at or above average, so a is not the Borda winner.

B10. In a Latin-square profile (with $m = n$) each alternative occupies each rank exactly once, so every rank sum equals $\binom{n+1}{2}$ and all Borda scores are equal. By the identity in B9, a Condorcet winner a satisfies $N(a, b) > m/2$ for all b, hence $s_B(a) > (n-1)m/2$, strictly *above* average — impossible when all scores are equal. So no Condorcet winner exists.

For $n = m = 3$ exhaustive enumeration of all $6^3 = 216$ profiles gives exactly 12 Latin squares, exactly 12 profiles with no Condorcet winner, and these two sets coincide. The coincidence is an artefact of small numbers. For $n = m = 4$ there are 576 Latin squares (all without a Condorcet winner, as the argument above requires) but 218,304 profiles without a Condorcet winner. An explicit non-Latin counterexample:

$$a \succ b \succ c \succ d, \quad b \succ c \succ a \succ d, \quad c \succ a \succ b \succ d, \quad a \succ b \succ c \succ d.$$

Here a beats b 3–1, b beats c 3–1, and a ties c 2–2, so no alternative beats all others; the rank column for a is $(1, 3, 2, 1)$, which repeats, so the array is not a Latin square.

B11. For each ordered pair (a, b) for which the target tournament has $a \to b$, introduce two voters: one ranking $a \succ b \succ c_1 \succ \cdots \succ c_{n-2}$ and one ranking $c_{n-2} \succ \cdots \succ c_1 \succ a \succ b$. On the pair $\{a, b\}$ both prefer a, contributing $+2$ to $\mu(a, b)$. On every other pair the two voters disagree, contributing 0. Doing this once per edge uses $2\binom{n}{2} = n(n-1)$ voters.

B12. Let S be the Smith set and suppose an optimal π ranks some $y \notin S$ above some $x \in S$. Scanning π downward from the topmost non-member, there is an adjacent pair with $y \notin S$ immediately above $x \in S$. Swapping them changes the Kemeny cost, by Proposition 8.2, only through the pair $\{x, y\}$, and the change is $N(y, x) - N(x, y) = -\mu(x, y) < 0$ because $x \in S$ and $y \notin S$ forces x to beat y pairwise. This contradicts optimality.

B13. $\sum_j d_K(\pi, \sigma_j)$ counts triples $(j, \{a, b\})$ on which π and σ_j disagree. Group by pair: for the ordered pair with $a \succ_\pi b$, the rankers who disagree are exactly those preferring b to a, of which there are $N(b, a)$. Summing over ordered pairs consistent with π gives the claim.

B14. By right-invariance (B1), substituting $\sigma \mapsto \sigma\sigma_0^{-1}$ is a bijection of $\mathfrak{S}_n$ preserving $d_K(\cdot, \sigma_0) \mapsto d_K(\cdot, \mathrm{id})$, so $Z(\theta) = \sum_\sigma q^{d_K(\sigma,\mathrm{id})} = [n]_q!$ with $q = e^{-\theta}$. For $n = 3$: $Z(\theta) = (1 + e^{-\theta})(1 + e^{-\theta} + e^{-2\theta}) = 1 + 2e^{-\theta} + 2e^{-2\theta} + e^{-3\theta}$.

B15. $\ell(\sigma_0, \theta) = -\theta \sum_j d_K(\sigma_j, \sigma_0) - m \log Z(\theta)$. Since Z is independent of σ_0 (B14), maximising in σ_0 means minimising $\sum_j d_K(\sigma_j, \sigma_0)$, which is the Kemeny objective. The hypothesis $\theta > 0$ is essential: at $\theta = 0$ the model is uniform and every σ_0 is an MLE, and for $\theta < 0$ the MLE would *maximise* total Kendall distance.

B16. Use the exponential-race representation. Let $E_a \sim \mathrm{Exp}(w_a)$ independently and rank alternatives by increasing E_a. The probability that a is first is $w_a / \sum_l w_l$, and by the memorylessness of the exponential the remaining order is generated by the same mechanism on the remaining items — which is exactly the Plackett–Luce construction. Hence the induced order on any subset is Plackett–Luce with the same weights, and in particular $\mathbb{P}[a \succ b] = \mathbb{P}[E_a < E_b] = \frac{w_a}{w_a + w_b}$.

B17. The rank of a is 1 plus the number of alternatives above it, so $\mathbb{E}[\sigma(a)] = 1 + \sum_{b \neq a} \mathbb{P}[b \succ a] = 1 + \sum_{b \neq a} \frac{w_b}{w_a + w_b}$ by B16. Each summand is strictly decreasing in w_a, so the expected rank is strictly decreasing in w_a and the Borda order recovers the weight order. For $w = (3, 2, 1)$ on (a, b, c): $\mathbb{E}[\sigma(a)] = 1 + \frac{2}{5} + \frac{1}{4} = 1.65$, $\mathbb{E}[\sigma(b)] = 1 + \frac{3}{5} + \frac{1}{3} = 1.9\overline{3}$, $\mathbb{E}[\sigma(c)] = 1 + \frac{3}{4} + \frac{2}{3} = 2.41\overline{6}$; the three sum to $6 = 1 + 2 + 3$, as they must.

B18. Let G act transitively and let P be G-invariant. Fix $g \in G$. Because g permutes the ballots of P among themselves, for every rank r the number of rankers placing a at rank r equals the number placing $g(a)$ at rank r. Hence a and $g(a)$ receive identical *multisets* of ranks, so $s_w(a) = \sum_j w_{\sigma_j(a)} = \sum_j w_{\sigma_j(g(a))} = s_w(g(a))$ for every weight vector w. Transitivity makes all scores equal. In the Latin-square case each alternative receives each rank exactly once, so the multisets are trivially identical and the conclusion holds for every positional rule without invoking G at all.

B19. The subgroups of $\mathfrak{S}_3$ are: $\{e\}$; three subgroups of order 2 generated by a transposition; $A_3 = \langle(1\,2\,3)\rangle$ of order 3; and $\mathfrak{S}_3$. Orbits: $\{e\}$ has three singleton orbits; $\langle(1\,2)\rangle$ has orbits $\{1, 2\}$ and $\{3\}$; A_3 and $\mathfrak{S}_3$ are transitive. Only A_3 and $\mathfrak{S}_3$ force a full three-way tie. A profile invariant under A_3 needs at least three ballots (e.g. P_2); one invariant under $\mathfrak{S}_3$ needs all six rankings, each with equal multiplicity. A profile invariant only under $\langle(1\,2)\rangle$ forces $s(1) = s(2)$ but leaves alternative 3 free — a partial, not total, obstruction.

B20. Base profile over $\{x, y, z_1, z_2\}$ with three rankers:

$$R_1 : x \succ y \succ z_1 \succ z_2, \quad R_2 : x \succ y \succ z_2 \succ z_1, \quad R_3 : y \succ z_1 \succ z_2 \succ x.$$

Then $\mathrm{RRF}_{60}(x) = 0.0484119 < \mathrm{RRF}_{60}(y) = 0.0486515$, so $y \succ x$. Now insert a new alternative w:

$$R_1 : x \succ w \succ y \succ z_1 \succ z_2, \quad R_2 : x \succ w \succ y \succ z_2 \succ z_1, \quad R_3 : y \succ z_1 \succ z_2 \succ x \succ w.$$

Now $\mathrm{RRF}_{60}(x) = 0.0484119$ (unchanged) and $\mathrm{RRF}_{60}(y) = 0.0481395$, so $x \succ y$. No ranker altered their opinion about x versus y: all three rank the pair the same way before and after ($x \succ y$, $x \succ y$, $y \succ x$). This violates Independence of Irrelevant Alternatives.

B21. $\sum_j d_K(\pi_F, \sigma_j) \leq \sum_j d_F(\pi_F, \sigma_j)$ uses the left half of Diaconis–Graham in the weak form $d_K \leq d_K + d_C \leq d_F$. The middle step $\sum_j d_F(\pi_F, \sigma_j) \leq \sum_j d_F(\pi_K, \sigma_j)$ is optimality of π_F for the footrule objective. The final step $\sum_j d_F(\pi_K, \sigma_j) \leq 2 \sum_j d_K(\pi_K, \sigma_j)$ is the right half.

B22. $\sum_a \sum_j (\pi(a) - \sigma_j(a))^2 = m \sum_a \pi(a)^2 - 2 \sum_a \pi(a) S(a) + \sum_a \sum_j \sigma_j(a)^2$ with $S(a) = \sum_j \sigma_j(a)$. The first and third terms do not depend on π (the first because π is a permutation of $[n]$), so we must maximise $\sum_a \pi(a)S(a)$. By the rearrangement inequality this is maximised when $\pi(a)$ and $S(a)$ are similarly ordered: the smallest position goes to the smallest rank sum. That is Borda.

B23. The constraints $x_{ab} + x_{ba} = 1$ with $x \in \{0, 1\}$ make the relation complete and antisymmetric, i.e. a tournament. A tournament is transitive if and only if it contains no directed 3-cycle: if it were not transitive there would be $a \to b$, $b \to c$ and not $a \to c$, hence $c \to a$, a 3-cycle. The constraint $x_{ab} + x_{bc} + x_{ca} \leq 2$ forbids exactly the configuration $x_{ab} = x_{bc} = x_{ca} = 1$, i.e. the 3-cycle. Hence the feasible set is precisely the set of total orders.

B24. $f(r) = -\log(k + r)$ has $f'(r) = -1/(k + r) < 0$ and $f''(r) = 1/(k + r)^2 > 0$: it is decreasing but *convex*, so $\sum_j f(r_j)$ is Schur-convex and, like RRF, is maximised by the most spread rank vector. Concavity is what is needed, not a heavier tail.

Take $f(r) = -r^2$ instead: $f' = -2r < 0$ and $f'' = -2 < 0$, so $\sum_j f(r_j)$ is Schur-concave. Equivalently, rank by *increasing* $\sum_j r_j^2$. On the two rank vectors: $(2, 2, 2)$ gives 12 and $(1, 1, 4)$ gives 18, so the consistent alternative wins — the reverse of RRF_{60}, which prefers $(1, 1, 4)$ by 2.48×10^{-5}.

Drawback: the squared-rank rule is dominated by deep ranks. A document at rank 500 in one list contributes 250,000, swamping everything else, so the rule is unusable without a truncation convention and a principled value for absent documents. RRF's bounded weights are precisely what make it robust here; the trade-off is real, and a sensible compromise is a bounded concave f such as $f(r) = -\min(r, N)^2/N^2$.

Set C — Hints and Discussion

C1. A sufficient condition follows from bounding the error term in Theorem 5.4. Distinct integer Borda sums differ by at least 1, so it suffices that $\frac{1}{k^2} > \left|\text{third-order remainder}\right|$; expanding gives a condition of the form $k \gtrsim mn$. Compare this with the simulation table: at $n = 10$, $m = 5$ the estimate suggests k of order 50, and agreement reaches 0.994 at $k = 60$.

C2. Write $X_{ab} = \sum_j(\mathbf{1}[a \succ_j b] - \mathbf{1}[b \succ_j a])$. Under impartial culture each X_{ab} has mean 0 and the triple (X_{ab}, X_{bc}, X_{ca}) is asymptotically trivariate normal. For a single voter and $u = \pm 1$ recording a versus b, $v = \pm 1$ recording b versus c, show $\mathbb{E}[uv] = \frac{1}{6} + \frac{1}{6} - \frac{1}{3} - \frac{1}{3} = -\frac{1}{3}$, so the limit is equicorrelated with $\rho = -1/3$. The orthant probability for a trivariate standard normal with equal correlations ρ is $\frac{1}{8} + \frac{3}{4\pi} \arcsin \rho$; doubling for the two cycle orientations gives $\frac{1}{4} - \frac{3}{2\pi} \arcsin \frac{1}{3} \approx 0.08774$, and Guilbaud's number is its complement.

C3. Straightforward simulation. The known asymptotic is that the probability of a Condorcet winner tends to 0 as $n \to \infty$ under impartial culture. Expect the cycle probability to pass $1/2$ somewhere around $n \approx$ 25–40; make sure your estimator has enough replicates to distinguish this from noise.

C4. Combine the two halves of B24: a bounded concave f with a defensible value for absent documents. Candidates: $f(r) = -\min(r, N)^2/N^2$, or $f(r) = \log \frac{N+1-r}{N}$ truncated. Evaluate on BEIR or MS MARCO with Recall@K at the reranker's input width (§10.6), not nDCG@5. Report whether any gain survives when the reranker is included.

C5. Treat each ranker as an "expert" with weight λ_j and each labelled query as an observation. The MM update is the Zermelo iteration of Chapter 9. Identifiability: λ is determined only up to positive scaling, so fix $\sum_j \lambda_j = 1$; also note that if two rankers produce identical lists their weights are individually unidentifiable.

C6. Try to adapt the standard Borda-to-Kemeny argument. The obstruction is that RRF is Borda *plus* a dispersion bonus (Theorem 5.6); construct families of profiles with many Borda ties and see whether the bonus can accumulate. A negative result would be as interesting as a positive one.

C7. Fagin, Kumar and Sivakumar (2003) define $d_{\mathrm{K}}^{(p)}$ by assigning penalty p to pairs where the relative order cannot be determined from both top-k lists. Show the triangle inequality holds up to a constant depending on p, and observe that p plays exactly the role of the zero-fill convention: it is the price of absence, made explicit.

C8. Baik–Deift–Johansson: for uniform $\sigma \in \mathfrak{S}_n$ the LIS satisfies $\ell(\sigma) \approx 2\sqrt{n}$ with fluctuations of order $n^{1/6}$ following the Tracy–Widom GUE law. Geometrically: almost all of $\mathfrak{S}_n$ sits at Ulam distance $n - 2\sqrt{n} + O(n^{1/6})$ from the identity, so $(\mathfrak{S}_n, d_{\mathrm{U}})$ is extremely concentrated near its diameter — Ulam distance carries almost no information for large n.

C9. $\mathbb{R}^{\mathfrak{S}_n}$ decomposes as $\bigoplus_\lambda (\dim S^\lambda) S^\lambda$ over partitions $\lambda \vdash n$. Ranked data analysis (Diaconis, *Group Representations in Probability and Statistics*, Ch. 8) shows the first-order marginals — the $n \times n$ matrix of "how often did a come rth" — span the $(n) \oplus (n-1, 1)$ isotypic components, and any positional rule is a linear functional on that space. So *every* positional rule, including RRF, sees only the $(n-1, 1)$ component; the choice of w merely selects a vector within it. This is the representation-theoretic form of Theorem 5.4: RRF and Borda differ only in which vector, and the difference vanishes as k grows.

C10. Vary the zero-fill value (from 0 to $1/(k+N)$ to $1/(k+N+1)$) while holding k, N, the retrievers and the queries fixed; separately vary k over two orders of magnitude while holding the zero-fill value fixed. Theorem 5.9 predicts the first sweep moves Recall@K substantially and the second barely at all. A finding that k matters more would falsify the claim.

C11. With $n = 2$ simple majority satisfies (U), (P) and (IIA) and is not a dictatorship. Lemma 7.4 breaks immediately: Steps 1 and 2 both require a third alternative c, which does not exist. The transitivity of $F(P)$ — the engine of the whole proof — is vacuous on two alternatives.

C12. The family of decisive coalitions is closed under supersets and finite intersections and contains exactly one of S, S^c for every S: an ultrafilter. On a finite set every ultrafilter is principal, generated by a single point — the dictator. On an infinite electorate non-principal ultrafilters exist (given choice), so non-dictatorial Arrovian rules exist but are non-constructive.

C13. Black's theorem: if preferences are single-peaked with respect to a common linear order on alternatives, the alternative at the median peak is a Condorcet winner and majority preference is transitive. For retrieval, single-peakedness would mean all retrievers order documents monotonically along one latent axis — plausible for a single narrow topic, implausible for mixed lexical/semantic queries. Testability: check whether the observed rankings are consistent with *some* common axis (this is a polynomial recognition problem).

C14. Use the ILP of Chapter 8 with a commercial or open solver; $n \le 40$ is comfortable. Expect the Kendall distance between Kemeny and RRF_{60} output to grow with pool size — and to be concentrated in the tail rather than the head, which is the practically reassuring finding.

Part VI

Appendices

CHAPTER A

Notation

A, n	set of alternatives; $n = \|A\|$
m	number of rankers / voters
$\mathfrak{S}_n$	symmetric group on n letters; the set of strict complete rankings
$\sigma(a)$	**rank vector**: the position of a under σ (small is good)
$\sigma^{-1}(r)$	**order vector**: the alternative in position r
$a \succ_\sigma b$	$\sigma(a) < \sigma(b)$
$P = (\sigma_1, \dots, \sigma_m)$	profile
$N(a, b)$	number of rankers preferring a to b
$\mu(a, b)$	majority margin $N(a, b) - N(b, a)$
$S(a)$	Borda rank-sum $\sum_j \sigma_j(a)$
$Q(a)$	second moment $\sum_j \sigma_j(a)^2$
$w = (w_1, \dots, w_n)$	positional weight vector, $w_1 \geq \cdots \geq w_n$
$\mathrm{RRF}_k(a)$	$\sum_j (k + \sigma_j(a))^{-1}$
d_{K}, d_{F}, d_{S}, d_{C}, d_{U}	Kendall, footrule, Spearman, Cayley, Ulam distances
τ, ρ	Kendall and Spearman correlation coefficients
$[n]_q!$	q-factorial $\prod_{i=1}^{n}(1 + q + \cdots + q^{i-1})$
$Z(\theta)$	Mallows normalising constant
$\mathbf{1}$	all-ones vector

CHAPTER B

Numerical Verification

Every numerical claim in this monograph was computed rather than recalled. The computations, all short, are:

1. **RRF weight tables** (Chapter 5): direct evaluation of $1/(k+r)$.
2. **Condorcet cycle probabilities** (Chapter 6): exhaustive enumeration over 6^m profiles for $m = 3, 5, 7$, giving the exact rationals $1/18$, $5/72$, $875/11664$.
3. **Guilbaud's number**: $\frac{3}{4} + \frac{3}{2\pi} \arcsin \frac{1}{3} = 0.9122601\ldots$
4. **Diaconis–Graham** (Chapter 3): exhaustive check over all 720 elements of $\mathfrak{S}_6$ against the identity. Both bounds hold; the left is tight 511 times, the right 132 times.
5. **RRF-to-Borda convergence** (Chapter 5): Monte Carlo over random profiles restricted to those with distinct Borda scores, 800 replicates per cell.
6. **Kemeny costs** (Chapters 5 and 8): exhaustive over all $n!$ orders for $n \leq 4$.
7. **Latin squares versus Condorcet cycles** (Problem B10): exhaustive over 216 profiles for $n = m = 3$ and 331,776 for $n = m = 4$, giving 12 and 576 Latin squares against 12 and 218,304 profiles without a Condorcet winner.
8. **RRF score resolution** (Chapter 5): enumeration of achievable fused scores for $m = 3$, $k = 60$ over the top 60 ranks: 36,791 distinct values, minimum positive gap 1.3×10^{-11}.
9. **The IIA counterexample** (Problem B20) and **the Condorcet failure of** RRF_{60} (Example 4.8): direct evaluation.

CHAPTER C

Reading and Examination Mapping

Primary sources

- K. J. Arrow, *Social Choice and Individual Values* (1951). The founding text.
- P. Diaconis, *Group Representations in Probability and Statistics* (1988), Chapters 5–8. The mathematical home of ranked data; the source for metrics on $\mathfrak{S}_n$, Mallows models and the representation-theoretic view.
- P. Diaconis and R. L. Graham, "Spearman's footrule as a measure of disarray", *JRSS-B* 39 (1977). The comparison inequality.
- C. Dwork, R. Kumar, M. Naor, D. Sivakumar, "Rank aggregation methods for the Web", WWW 2001. NP-hardness, the footrule 2-approximation, and the beginning of the computational branch.
- G. V. Cormack, C. L. A. Clarke, S. Büttcher, "Reciprocal rank fusion outperforms Condorcet and individual rank learning methods", SIGIR 2009. The origin of $k = 60$.
- H. P. Young, "Social choice scoring functions", *SIAM J. Appl. Math.* 28 (1975); H. P. Young and A. Levenglick (1978). The two characterisation theorems.
- R. Fagin, R. Kumar, D. Sivakumar, "Comparing top-k lists", SODA 2003.

Mapping to Indian postgraduate examinations

Syllabus topic	Where in this book	Problems
Permutation groups, cycle type, conjugacy	§2.1, §3.3	A1, B1, B19
Group actions, orbits, transitivity	§2.4, §5.7	B18, B19
Counting: Stirling numbers, ordered Bell	§2.3	A7, A15
Generating functions, q-analogues	§3.1	A16, A17
Metric spaces: axioms, verification	Ch. 3	B1–B3
Rearrangement inequality	§3.4	B22
Jensen, convexity, majorisation	§5.4	B7, B24
Elementary MLE	Ch. 9	B15, B16, B17
Linear programming, integer programming	§8.2	B23
Asymptotic expansions	§5.3	B8
CLT and orthant probabilities	§6.2	C2

CHAPTER D

Quick Reference

Reciprocal Rank Fusion. $\mathrm{RRF}_k(a) = \sum_j \frac{1}{k + \sigma_j(a)}$, default $k = 60$; absent documents contribute 0.

Five things that are true.

1. It is a positional scoring rule with $w_r = 1/(k + r)$ — nothing more exotic.
2. On short lists at $k = 60$ it *is* Borda (Thm. 5.4).
3. At equal rank-sum it prefers the *polarising* item, not the consistent one (Thm. 5.6).
4. The consensus effect practitioners observe comes from zero-filling absent documents: at $m = 3$, $k = 60$, rank 122 everywhere beats rank 1 once (Thm. 5.9).
5. Its real virtue is scale-freeness: no normalisation, no calibration, no training.

Three things it cannot do.

1. Satisfy IIA — adding a document can flip two others (Arrow).
2. Elect the Condorcet winner reliably (Thm. 4.7).
3. Break a symmetric profile — and neither can anything else (Thm. 5.12).

Two diagnostics worth logging. Consecutive score gap below $\frac{1}{(k+1)(k+2)} = 2.64 \times 10^{-4}$ at $k = 60$; and a Smith set of size greater than one on the top K.

One thing to do instead of tuning k. Rerank.