

NATURAL INTELLIGENCE



THE AICE LIBRARY

Natural Intelligence

*A field manual of human wit – puzzles, lateral thinking, and the
arts of reasoning machines haven't mastered*



FIRST EDITION · JULY 2026

NATURAL INTELLIGENCE — A Field Manual of Human Wit

Composed by the AICE team. English edition, July 2026.

This book was made the way it teaches. The fifty-puzzle canon carries corrected attributions, stated in the text rather than hidden. Sixteen solutions were machine-verified by exhaustive enumeration or simulation before printing. Every claim about artificial intelligence is dated — July 2026 — because those sentences are built to age honestly.

One riddle is posed at the front of this book and answered only in the colophon. Readers who skip ahead get the answer and lose the point.

Set in Lora and Work Sans. Built with WeasyPrint. A book of the AICE Library — books/thinking/natural-intelligence.

*"What goes on four legs in the morning, two legs at noon,
and three legs in the evening?"*

THE SPHINX — KEEPING HER GATE SINCE THEBES

The answer is on the last page of this book. Everything between here
and there is training for the gates she keeps now.

Contents

How to Use This Book	11
Introduction — The Case for Wit	15
THE LINEAGE	19
1. Riddles at the Gate	21
2. The Golden Age	31
3. Thinking About Thinking	41
THE MOVES	49
4. Breaking the Frame	51
5. Pure Deduction	61
6. Numbers, Chance, and Shape	71
7. The Boss Room	85
THE OLYMPIAD MIND	97
8. How Champions Think	99
9. The Kill Shots	109
10. The Builder's Tools	123
11. The Modeler's Eye	137
THE EDGE	151
12. The Scoreboard	153
13. The Moats	163
14. The Race	173
THE GYMNASIUM	183
15. Training the Edge	185
16. The Thirty-Day Regimen	195
Appendix C — The Reserve Rack	205
Appendix A — The Moves Index	217
Appendix B — The Graveyard of Thinking Myths	221

CONTENTS

Further Reading	225
Colophon	227

How to Use This Book

This book asks you to do something books rarely ask: stop reading. More than fifty times in these pages, a puzzle will be set before you, and everything the book can teach lives in the minutes you spend alone with it. Read straight through and you will collect trivia. Stop, struggle, and the same pages become training.

The ritual

Every puzzle here follows one format. First the statement, with every number you need. Then a single italic line — *pause here, the page can wait* — which is the last exit before the answer (its wording varies; its meaning never does). The solution follows immediately, on the same page, because a solution you must hunt for is a solution you'll skip. So cover it. Keep an index card as a bookmark and use it as a shutter. Wrestle with the statement until you solve it, or until you hold something better than nothing: a plan that failed, a pattern you suspect, a reason the obvious answer can't be right. Then read on. A failed attempt converts the solution from information into correction, and correction is where learning happens.

How long to hold out: five or ten minutes for the easy ones; give a hard one a whole commute, or a night. The one unforgivable move is peeking at thirty seconds. Being stuck is not a malfunction. Being stuck is the product.

The stars

- ★ — anyone can solve it: one honest idea, no special knowledge.
- ★★ — an engaged amateur gets there.
- ★★★ — a solid challenge; budget a false start or two.
- ★★★★ — hard; most solvers want a hint or a night's sleep.
- ★★★★★ — killers; solving one unaided is an event worth telling someone about.

Stars measure the distance to the idea, not the arithmetic.

The Moves

Every solution ends the same way: **The Move**, one paragraph naming the reusable pattern the puzzle just taught — find the invariant, condition on the procedure, solve the endgame first. The Moves, not the puzzles, are the curriculum; the puzzles are delivery vehicles. There are about thirty Moves in all, indexed in Appendix A, and they surface in debugging, negotiation, medicine, and law far more often than in parlor games. When one fires in your daily life, say its name. Naming is how a trick becomes a tool.

The parts

Part I is the lineage: close to four thousand years of the puzzle, from a scribe's joke to a machine's entrance exam. **Part II** is the armory — the core Moves, taught puzzle by puzzle, easiest first. **Part III** is the same art under pressure: how olympiad solvers actually think. **Part IV** is the honest scoreboard of human wit against artificial intelligence, every claim dated. **Part V** is the gymnasium: a thirty-day regimen, and what the evidence on training can and cannot promise.

Three ways through

- **The puzzler's path.** Chapter 1 first — your opening ritual waits there — then Parts II and III in order, with the history chapters as rest days and Part V when you're ready for a program.
- **The scholar's path.** Front to back. The book is built in that order: where the art came from, what it is, who is coming for it, how to keep it.
- **The manager's path.** Short on time: the Introduction, then Chapter 8, then the Move recaps at each chapter's end plus Appendix A, then Part IV. Solve one ★★ puzzle with your coffee; the habit outperforms the reading.

Small print

Claims that carry real weight are graded inline: ⚙ mechanism · 🔒 solid evidence · ♦ useful heuristic · ⚠ contested. In Part IV, capabilities also wear tags

— DURABLE , ERODING , FALLEN — and every AI claim carries its date, because that field moves monthly and a July 2026 statement is a July 2026 statement. Dates are CE throughout. Corrected history is stated plainly: some famous attributions in this genre are false, and this book names them.

One more thing. The first riddle in this book — the oldest gate-keeping riddle we have — is posed in Chapter 1 and answered nowhere in the sixteen chapters that follow. Its answer waits in the colophon, on nearly the last page. If you can't stand the wait, there is a remedy: solve it.

Introduction — The Case for Wit

Around 1550 BCE, an Egyptian scribe named Ahmes sat down to copy a mathematical scroll that was, by his own note, already some two centuries old. Most of it is sober professional matter — grain, bread, the geometry of slopes. Then comes Problem 79. Seven houses. In each house, seven cats. Each cat kills seven mice; each mouse would have eaten seven heads of spelt; each head of spelt would have yielded seven hekats of grain. Ahmes climbs the ladder — 7, 49, 343, 2,401, 16,807 — and writes the total: 19,607.

No granary needed that number. Nobody's taxes depended on cats. Problem 79 is arithmetic for the joy of it — a geometric series wearing a costume — and it is one of the oldest pieces of recreational thought that survives.

It also refuses to die. Nearly three thousand years after Ahmes, Leonardo of Pisa put the same joke in his *Liber Abaci*: seven old women on the road to Rome, each with seven mules, each mule with seven sacks. Five centuries after that, English children were chanting it — *as I was going to St Ives, I met a man with seven wives*. A good puzzle, it turns out, is a meme with a three-thousand-year half-life.

This book is about that thread — close to four thousand years of it — and about what the thread has been doing to us all along.

The oldest gymnasium

A puzzle is a thinking pattern wearing a costume. Strip the cats and the spelt from Problem 79 and you hold exponential growth. Strip the river and the boat from a certain eighth-century brainteaser and you hold state-space search. Strip the doors from a certain game show and you hold conditional probability — the real kind, the kind that depends on how the evidence reached you. The costumes date; the patterns don't. A dozen or so deep structures have been re-dressed and re-sold across millennia — the magic square becomes sudoku; the riddle at the gate becomes an entrance exam for machines — because the muscles they train keep mattering.

Around 800 CE, someone at Charlemagne's court — tradition says Alcuin of York — collected fifty-odd problems under a title that gives the whole game away: *Propositiones ad acuendos juvenes*. Problems to sharpen the young. Not amuse: sharpen. That is the oldest explicit statement of this book's thesis: puzzles are the oldest gymnasium of the mind. Every era installed new equipment. Magic squares taught constraint and invariant. River crossings taught planning. Weighing puzzles taught the economics of information. Ciphers taught that structure leaks even when content is hidden. The gymnasium grew for forty centuries, one apparatus at a time, and nearly all of it still works.

The capacity all that equipment trains deserves a plain name, and this book uses an old one: **wit** — the fast, flexible, resourceful end of intelligence; the part that notices, reframes, and pounces. Not accumulated knowledge, and not calculation. The thing that moves when the problem is new.

Why a manual, and why now

For almost all of those forty centuries, "the best available puzzle-solver" meant a person, by definition. That stopped being automatic very recently — within the lifespan of the book you are holding. Machines now solve, at or beyond strong-human level, a good share of the closed, well-specified problems this tradition spent millennia polishing. Some trophies everyone assumed were safe have fallen. Others — this is the interesting part — have barely moved. Which fell, which stand, and what separates the two is the business of Part IV, where every claim carries a date and a source, because in that arena a sentence's truth has a shelf life.

One thing is worth saying in advance. The capacities that remain most stubbornly human are not exotic. They are the ordinary-sounding ones: noticing what kind of problem you are actually in; abandoning a bad representation for a better one; knowing what you don't know and saying so; owning a goal across a long, messy stretch of world. Those happen to be, almost exactly, the muscles the puzzle tradition has always trained. That coincidence is the case for this book. Natural intelligence spent four thousand years building its own gymnasium; now, with a second species of solver on the premises, seems a poor moment to cancel the membership.

What this book promises

Four things.

Verified play. The puzzles here are stated exactly and solved fully, and the set-piece solutions were checked by machine — enumerated, simulated, or searched exhaustively — before the prose was written around them; the colophon has the details. Where certainty runs out, the text says so. One famous paradox in these pages is presented as what it genuinely is: open.

Honest history. This genre's past includes theft and fabrication, and the record here is stated plainly. The most famous puzzle fraud in American history is told in full in Chapter 2. The "Einstein's riddle" that circulates online was written by no Einstein. Where a beloved anecdote is embellished lore, it is labeled lore. Puzzling is a game of exact truth; its history deserves the same standard.

The Moves. Every puzzle ends by naming the reusable pattern it teaches. About thirty such Moves cover the whole canon, and they — not the puzzles — are the payload. They are what transfer to the debugging session, the diagnosis, the negotiation. Appendix A indexes every one.

A program. Part V turns all of it into training you can actually run — a thirty-day regimen, a maintenance dose, and honest measures of progress.

What this book refuses

No mystique, and no snake oil. The commercial brain-training industry built itself on the claim that puzzle practice makes you *generally* smarter; the best evidence says trained tasks improve trained tasks, and the miracle largely stops there $\triangle$ — Appendix B holds a full graveyard of thinking myths, that one included. This book claims something narrower and more defensible: specific, nameable patterns, learnable from puzzles, that transfer as far as you deliberately carry them — and a clear-eyed account of which ones matter more, not less, as machines improve.

It also refuses manufactured difficulty. You will meet no "only 2% of people can solve this" in these pages; that number was always marketing, stapled to a magazine puzzle decades after the fact. Difficulty here is graded in stars and earned honestly. If something is easy, it says so. If something took mathematics a century to settle, it says that too.

The wager

Wit has always been tested at gates. The next chapter opens at the most famous gate of all, where a monster with a lion's body sat beside the only road into a city and asked travelers one question. Wrong answers were fatal. The question is still worth asking — you will find it on the first page of Chapter 1, and you will not find its answer until the colophon, partly as a joke, mostly as a promise: the point of a riddle is the state it puts you in, not the syllable that discharges it. Four thousand years of minds have loved that state enough to keep building new gates for the pleasure of standing at them.

The gate is open. Come be sorted.

PART I

The Lineage

CHAPTER 1

Riddles at the Gate

Outside Thebes, on a rock beside the only road into the city, sat a monster: lion's body, eagle's wings, a woman's face. She stopped every traveler and asked one question. Answer wrong, and she killed you. The city starved behind her, because the riddle at the gate had closed the road.

The question was this:

What goes on four legs in the morning, on two legs at noon, and on three legs in the evening?

By the time the story was written down — the riddle's exact words reach us only through later authors, Apollodorus among them — Greece had already decided what a riddle was for. Not entertainment. Sorting. The Sphinx is a filter with a face: one question, engineered so that a literal mind slides straight off it, with everything staked on the answer. Every noun in it is a trap. And when a traveler named Oedipus finally answered, the story says, the Sphinx flung herself from the rock and died. A riddle spent is a riddle dead.

This book will not spend it here. The answer waits in the colophon, on nearly the last page, where accounts are settled. If you don't already know it, don't look it up — you now own something rarer than a solution: a live riddle, the oldest one there is. Sit with the itch. The itch is this book's subject.

What matters for this chapter is the shape of the scene: a mind, a gate, and a question standing between them. That is the oldest job puzzles do. Long before they were pastimes, they were tests of worth. What follows walks the lineage's first twenty-seven centuries — a scribe's joke, a lethal argument, a turtle's shell, a general's headcount, a monk's workbook, a mer-

chant's textbook — and watches each one bolt a permanent piece of equipment into the gymnasium of the mind.

The Scribe's Joke

The oldest puzzle we can still read survived by the grace of dry sand. Around 1550 BCE an Egyptian scribe named Ahmes copied out a mathematical scroll — itself, he notes, taken from a text some two centuries older — filled with sober professional material: grain shares, bread and beer arithmetic, the slopes of pyramids. And then there is Problem 79, which is sober about nothing. Seven houses. In each house, seven cats. Each cat kills seven mice. Each mouse would have eaten seven heads of spelt; each head of spelt would have yielded seven hekats of grain. Ahmes runs up the ladder — 7, 49, 343, 2,401, 16,807 — and totals it: 19,607.

No granary needed that number. There is no administrative reason on earth to multiply cats. Problem 79 is a geometric series wearing a costume, arithmetic performed for the joy of the climb, and it establishes something worth pausing on: abstraction-as-play is as old as writing itself. From the moment humans possessed a tool for thinking, some of them immediately began playing with it — and the play, not the grain accounts, is what we are still solving thirty-six centuries later.

Hold on to Ahmes and his sevens. They will return before this chapter ends.

The Paradox Engine

Around 450 BCE, in the Greek town of Elea, Zeno built puzzles of a more menacing kind: arguments. Achilles, fastest of men, grants a tortoise a head start. By the time he reaches the tortoise's starting point, it has crawled ahead; by the time he covers that gap, it has crawled again; and so on, without end. Each step of the argument is irresistible. The conclusion — the fastest runner can never pass the slowest — is absurd. Zeno's dichotomy is worse still: to cross a room you must first cross half of it, and before that half of the half; motion cannot even begin.

A riddle kills you at the gate. A paradox moves into your house. Zeno's arguments shrugged off every commonsense rebuttal for roughly 2,300 years,

until mathematics built limits and convergent series in earnest — the machinery by which an infinite pile of steps can sum to one finite walk. Absorb that number for a moment: the puzzle outlived the civilization that posed it, outlived a dozen civilizations after that, and sat waiting, politely unanswerable, for calculus to be invented. And that is the lesson a good paradox teaches, and it has nothing to do with tortoises: a paradox that will not die marks, with great precision, the place where a concept is missing. The discomfort is a map reference, not a malfunction. This book will meet paradoxes doing the same cartography in probability and in logic, and it will treat each one the same way: not as a nuisance, but as a survey stake driven into unmapped ground.

The Turtle's Gift

Chinese legend says a turtle crawled out of the River Luo bearing a pattern of dots on its shell. However it actually began — the legend is ancient; the earliest surviving texts recording the square date to the Han era — the object itself is nine numbers in a three-by-three grid: 4, 9, 2 across the top; 3, 5, 7 across the middle; 8, 1, 6 across the bottom. Every row, every column, and both diagonals sum to 15. Each digit from 1 to 9 appears exactly once. This is the Lo Shu, the ancestral magic square, and it is essentially unique — every 3×3 magic square using 1 through 9 is this one, rotated or reflected.

The surface pleasure is constraint satisfaction: nine numbers, eight lines, one arrangement obeying all of them at once. But hiding inside is a deeper idea. The 15 is not a decoration; it is a necessity. The nine digits total 45, and the three rows must split that total equally — so every line was forced to 15 before a single digit was placed. Find the quantity that must hold regardless of arrangement, and you know things about arrangements you have never examined. This is the lineage's first **invariant** — the single most transferable idea in this book, and this chapter will not be the last place it draws blood.

The Lo Shu's descendants run twelve centuries deep, sudoku among them; Chapter 3 picks up that story.

The Old Man's Toys

Archimedes of Syracuse spent his gravest hours on war machines and, the record suggests, his happiest ones on toys. Two survive, and both run deeper than almost anything else in antiquity.

The first, the *Stomachion*: fourteen flat pieces that assemble into a square. For two thousand years it looked like a child's amusement. Then a re-reading of the Archimedes Palimpsest — the medieval prayer book whose scraped parchment still faintly carries his treatise — suggested what he was actually asking: in how many ways can the fourteen pieces form the square? Not *whether*. *How many ways*. Counting configurations, treated as mathematics — combinatorics, two millennia early. The answer took until 2003, when Bill Cutler settled it by computer: 17,152.

The second toy arrives as a poem: the Cattle Problem, an epigram sent to Eratosthenes, librarian of Alexandria. Compute the size of the Herd of the Sun, it begins, and then unspools a cascade of innocent-sounding conditions on white, black, dappled, and brown bulls and cows. The conditions compress into what we now call a Pell equation, and the smallest herd satisfying them all is a number 206,545 digits long. It was not written out in full until 1965, by machine. A pasture problem, posed as a taunt between scholars, holds more cattle than the observable universe holds atoms.

Archimedes' twin lessons are the permanent equipment here: counting arrangements is real mathematics, and a trivially posed puzzle can be astronomically deep. Simple surface tells you nothing about depth — a warning this book will repeat in every century.

One postscript belongs to the same family. The tangram — seven pieces, one square, endless silhouettes — radiates ancient-Chinese aura, and popular histories long gave it a millennial pedigree. Honesty compels the correction: its documented trail begins only around 1800, in China, from which it crossed to Europe and America within two decades and set off one of the first international puzzle crazes. Dissection play thus brackets this whole history — genuinely ancient in Syracuse, freshly minted in Qing-era China — and the craving it feeds never changed: the same pieces, seen differently, become something else.

The Remainder Trick

A Chinese arithmetic text of the fourth or fifth century CE, the *Sunzi Suanjing*, poses what sounds like small-town bookkeeping: there are things whose number we do not know; counted by threes they leave two; counted by fives they leave three; counted by sevens they leave two. How many things? The smallest answer is 23. Legend attaches the method to the general Han Xin, counting troops by having them form ranks of three, five, and seven and reporting only the remainders — the count kept secret even from the counted.

The move underneath is the treasure: when the whole is unknowable, split it into knowable residues and reassemble. That instinct grew into a theorem — the Chinese Remainder Theorem, as the world now calls it — and the theorem grew into infrastructure: modular decomposition runs today inside the cryptography that moves your money. It will not be the last time in this lineage that a riddle grows up to be load-bearing.

The Monk's Workbook

Around 800 CE, at the court of Charlemagne, the emperor's chief scholar — tradition names Alcuin of York, and the attribution, though not airtight, is old — assembled a Latin collection of arithmetic problems: fifty-three in most manuscripts, fifty-six in some. Its title is a manifesto. *Propositiones ad acuendos juvenes*: Problems to Sharpen the Young. Not to amuse them. To sharpen them. It is the first collection in history to say out loud what every setter since Ahmes had known: these things are training equipment.

And the training it introduced was genuinely new. The manuscript's river-crossing problems are the first state-space planning problems on record — a world of discrete situations, legal moves, and forbidden configurations — posed eleven and a half centuries before artificial-intelligence researchers would define "search" and rediscover the genre as their own foundation. Notice what kind of difficulty this is. Nothing in these problems requires mathematics beyond a child's; everything in them requires *thought*. The hardness lives entirely in the representation — in seeing the situation as a

set of states and the constraints as walls — which is exactly why a piece of Carolingian homework still feels modern.

Problem 18 of the collection is the most famous puzzle in this book. It has been sharpening the young for twelve hundred years, and it is your first ritual.

★ 1 • Wolf, Goat, Cabbage

A farmer must ferry a wolf, a goat, and a cabbage across a river. The boat carries only the farmer plus one passenger — wolf, goat, or cabbage. Left alone together, the wolf eats the goat, and the goat eats the cabbage. How does everything reach the far bank intact?

Pause here. That river has waited twelve centuries; it can wait ten minutes more.

Solution. The first move is forced, though it rarely feels that way. Ferry the wolf first, and the goat, left behind, eats the cabbage. Ferry the cabbage first, and the wolf eats the goat. Only one departure leaves peace on the shore: **take the goat** (crossing 1) — wolf and cabbage tolerate each other fine. Row back empty (2). Now take either remaining passenger; say the wolf (3). And here the puzzle bares its teeth: landing the wolf beside the goat serves lunch — unless you do the thing that feels like failure and **bring the goat back with you** (4). Now the far bank holds the wolf, alone and safe; the near bank holds you, the goat, and the cabbage. Leave the goat; carry the cabbage over (5) — wolf and cabbage, peaceful again. Row back empty (6). Fetch the goat (7).

Done in seven crossings — and seven is the minimum; an exhaustive search of every possible plan confirms it, and finds exactly one other optimal route: the mirror image, with wolf and cabbage trading order. Why must the goat shuttle? Look at the cast. The wolf conflicts only with the goat. The cabbage conflicts only with the goat. The goat conflicts with both — it eats and is eaten. It is the bottleneck, the one passenger that can never be left with anything on either bank, so it is the one that must commute. And crossing 4, the move that cracks the puzzle, is the move beginners refuse to

consider, because it looks like undoing progress. It is not. Progress here is measured in the space of situations, not in cargo delivered — and the shortest path through that space runs, for one crossing, backward.

The Move — Shuttle the bottleneck. Identify the element that conflicts with everything, and let it commute; accept moves that look like retreat. Any plan graded by states rather than by visible progress works this way — the refactor that makes the code briefly uglier, the negotiation that reopens a settled clause to unlock two others. Backward is a direction, not a verdict.

The rest of the workbook plays variations: jealous husbands who will not leave their wives in other men's company, bigger boats, tighter constraints — the same machinery in a thicker web of states. And notice what the title promised and delivered. No student of Alcuin's was ever going to ferry an actual wolf. The wolves were never the point; the planning muscle was the point. Twelve hundred years later, when an engineer teaches a warehouse robot to sequence its moves without ever entering a forbidden state, she is working problem 18 at scale.

The Salesman of Numerals

In 1202, a young man from Pisa published a mathematics textbook with a sales problem. Leonardo — called Fibonacci — had grown up around his father's trading posts in North Africa, where he had learned the Hindu-Arabic numerals: nine digits and a zero, positional, built for calculation. Europe, meanwhile, was doing long division in Roman numerals, which is a form of penance. His *Liber Abaci* had to persuade merchants to abandon the notation of their fathers.

The case wrote itself, for anyone willing to attempt long multiplication with an L, two Xs, and a handful of Is; the trouble was inertia, guild habit, and distrust of a zero that seemed to mean nothing at all. So his instrument of persuasion, in large part: puzzles. The famous one asks how rabbit pairs multiply — each adult pair breeding a new pair monthly, each newborn pair maturing in a month — and the census month by month runs 1, 1, 2, 3, 5, 8,

13, each term the sum of the previous two. Growth, captured by a recurrence: the sequence that now carries his name through mathematics and half the sciences began life as a word problem about rabbits. And elsewhere in his exercises sits an old friend: seven old women on the road to Rome, each with seven mules; each mule carries seven sacks; each sack, seven loaves. Ahmes' ladder of sevens, twenty-seven centuries on, still climbing. Five more centuries and it becomes an English rhyme about a man met on the road to St Ives. A good puzzle is a meme with a three-thousand-year half-life.

Fibonacci's addition to the gymnasium is quietly enormous: notation is a thinking tool, and puzzles are its marketing department. People do not switch representations because you lecture them; they switch because the new representation makes delightful things easy. The numerals won. The rabbits helped.

The Gate, Multiplied

Count the equipment installed across these twenty-seven centuries. Series-play, from a scribe with a sense of humor. The paradox, a flare fired over missing concepts. The invariant, a turtle's gift. Deep counting, from an old man's toys. Modular decomposition, from a general who would not share his headcount. State-space planning, from a monk who wrote "sharpen" and meant it. The recurrence, sold by rabbits. None of it has expired. Every one of these moves will be asked of you again, in costume, before this book ends.

Ahead, the lineage leaves the manuscript for the printing press and goes industrial — bylines, crazes, prize money, and the genre's first great fraud. A postmaster's wooden toy will paralyze two continents for a season; a thousand-dollar prize will sit in perfect safety behind a wall of arithmetic; and the era's loudest genius will turn out to have stolen his most famous credit. One signpost for the longer road: five centuries after Fibonacci, in 1736, a stroller's idle question about crossing the seven bridges of Königsberg without repeating one would hand Leonhard Euler the founding move of modern problem-solving — throw away everything except structure — and create graph theory in the bargain; that story is told in full in Chapter 11.

The Sphinx, meanwhile, is still on her rock, and her question is still open on your table. Morning, noon, evening; four, two, three. There is a whole book of training between here and the colophon. Use it.

CHAPTER 2

The Golden Age

For its first three thousand years, the puzzle traveled hand to hand — papyrus to scroll, scroll to monastery vellum — at the pace of ink and luck. Then Europe built the printing press, and later the steam press, the penny newspaper, and the commuter railway, and the puzzle discovered what it could do with an audience of millions. Between 1612 and 1945 the genre acquired everything we now think of as modern: books and bylines, columns and crazes, prize money and plagiarism, and at last scientific respectability. It also staged two instructive spectacles — a fraud that fooled the world for a century, and a mass demonstration that effort cannot beat arithmetic. This chapter is the story of that age, told through the people who built it.

The Bookshelf

It begins, properly, with a book. In 1612 Claude-Gaspard Bachet published *Problèmes plaisants et délectables* — "pleasant and delightful problems" — the first true book of recreational mathematics: number tricks, Josephus-style elimination games, and a deceptively humble question about weights. With just four weights — 1, 3, 9, and 27 pounds — and a two-pan balance on which weights may sit in either pan, you can weigh every whole-pound load from 1 to 40. Inside that party trick hides ternary arithmetic, smuggled into the reader's head as shopkeeping. Bachet's deeper gift was canon-formation: puzzles now had provenance — who posed what, and when — which is precisely what the coming age would learn to fake. (The same Bachet edited the edition of Diophantus in whose margin Fermat would scribble the most famous note in mathematics.)

In 1694 Jacques Ozanam followed with *Récréations mathématiques et physiques*, four volumes — the encyclopedia stage, the reference that fed every nineteenth-century puzzle writer. The armory was stocked and

shelved. What the genre lacked was a detonator, and industrial printing supplied it: cheap paper, mass literacy, and dailies hungry for anything that made a reader buy tomorrow's edition.

In the winter of 1880, the world found out what that combination could do.

The Craze

Around 1874, Noyes Palmer Chapman, the postmaster of Canastota, New York, built a small wooden tray holding fifteen numbered sliding blocks and one empty space. It traveled quietly for a few years — friends, family, a few copies — and then, in January 1880, it went off like a charge. For four months, roughly January to April, the 15 Puzzle was a world craze: pushed in parlors, offices, and lecture halls across America and Europe, denounced by editorialists, blamed for lost workdays. Chapman filed for a patent only that March, as the wave was already cresting.

The engine of the mania was a specific challenge. Set up the board in order, but with tiles 14 and 15 swapped; slide until everything reads correctly. A thousand-dollar prize was attached to it — an offer whose loudest and longest promoter was the American puzzle showman Sam Loyd, a name to hold on to. Millions pushed tiles. Fortunes of attention were spent. And the entire time, the door was already locked: in 1879–80, in the *American Journal of Mathematics*, two mathematicians named Johnson and Story proved that the challenge is impossible. Not unsolved — impossible, the way five can't be even.

It is the book's next ritual, and it earns its three stars honestly: the puzzle is not to find a solution but to see, with certainty, why none exists.

★★★ 38 • The 14-15 Puzzle

A 4×4 tray holds fifteen sliding tiles, numbered 1–15, and one empty cell. A move slides an adjacent tile into the empty cell. In the solved board, the tiles read 1 through 15 in order, left to right, top to bottom, with the empty cell last, in the bottom-right corner. You are handed the board *almost* solved: everything is in place except that 14 and 15 have traded positions. By

sliding alone, restore perfect order. Sam Loyd offered \$1,000 for a solution. Should you try?

Pause here. Decide, at least, what kind of evidence would settle the question.

Solution. Don't try. No sequence of moves works — and the proof never examines a single sequence.

Count the empty cell as a phantom sixteenth tile. Any position of the board is then an arrangement of sixteen pieces in sixteen cells, and every move is one *swap*: the blank trades places with a neighboring tile. Now track two bookkeeping quantities.

First, the arrangement's parity. A basic fact about arrangements: a single swap of two pieces always flips an arrangement between "even" and "odd" — where even means reachable from the solved order by an even number of swaps, odd means odd. (Order matters not; the count's parity is fixed.) Since every move is exactly one swap, *the arrangement's parity flips on every move.*

Second, the blank's distance from home, counted in grid steps (rows plus columns away from the bottom-right corner). Every move drags the blank one step up, down, left, or right, so this distance changes by exactly ± 1 each move: *its parity also flips on every move.*

Two lights, blinking in perfect sync. So consider their **sum**. Each move flips both parities at once, and flipping both leaves the sum's parity unchanged — the sum modulo 2 is an *invariant*: no legal move, and therefore no sequence of legal moves however long or ingenious, can ever alter it.

Now evaluate the two boards. The solved board: zero swaps from itself (even), blank at home, distance zero (even). Even plus even — the invariant reads *even*. The prize board: exactly one swap from solved — 14 and 15 traded — so its arrangement is odd, while its blank sits at home (even). Odd plus even — the invariant reads *odd*. Even can never become odd under moves that preserve the sum. The prize position is unreachable. Every one of the millions of solvers was pushing tiles around one half of a universe whose other half they could never enter: exactly half of all arrangements

are reachable from the solved board — $16!/2$ of them, about ten trillion — and the theorem checks exhaustively in miniature: on a toy 2×2 tray, exactly 12 of the 24 arrangements can be reached.

Loyd's thousand dollars was perfectly safe. He knew it.

The Move — Impossibility by invariant. Before grinding at any transformation problem — tiles, a twisted cube, a database migration — hunt for a quantity that every legal move conserves, and check whether start and goal agree on it. If they disagree, you are done, gloriously: ten minutes of invariant-hunting can refute ten thousand hours of trying. This is the Lo Shu's forced 15, promoted from arithmetic to dynamics, and it may be the single most valuable reflex in this book.

The public of 1880 did not have the proof, and pushed on for months — the first mass collision between honest effort and an invariant, and effort lost, as it always does. The lesson was expensive, was not learned, and will be offered again: more than one of this book's hardest puzzles is solved by refusing to start.

The Showman

Sam Loyd was the greatest puzzle showman who ever lived, and that sentence needs both halves. An American already famous in his teens for chess problems, he had a genius for the puzzle as *spectacle*. He sold his Trick Donkeys puzzle to P. T. Barnum — cut three cards, rearrange, and two riders somehow mount three donkeys. His 1896 masterpiece, *Get Off the Earth*, put thirteen warriors on a rotating disc; turn it slightly and count again — twelve. No warrior is hidden. Each figure has quietly donated a sliver of itself to the others, and the thirteenth exists only as the sum of slivers. The vanishing warrior still teaches the right question — *watch what is actually conserved* — and Loyd's posthumous *Cyclopedia* (1914) piled up a mountain of such material, not all of it his: he lifted work freely, including from the English puzzlist you are about to meet.

Then there is the fraud. From 1891 until his death, Loyd claimed — in print, in interviews, decade after decade — that he had invented the 15 Puzzle and

engineered the craze of 1880. The claim entered the reference books and sat there, repeated in good faith, for over a century. It was demolished only in 2006, when Jerry Slocum and Dic Sonneveld published an archival investigation, *The 15 Puzzle Book*, tracing the puzzle's actual paper trail — to the postmaster of Canastota, Noyes Palmer Chapman. Loyd had not even entered the story until the mania was old news. Note the chronology with appropriate awe: the impossibility proof (1879–80) predates Loyd's claim of invention (1891). He spent his fame promoting a prize he knew mathematics had already voided, for a puzzle he did not create.

Genius and fraud in one biography — and, for the reader, a permanent calibration: this genre's two recurring cons are **false attribution** and **fake difficulty**, and they are as old as the business itself. You have now seen the first in its purest form; the "Einstein's riddle" nonsense of a later chapter is the same trick with a borrowed corpse. Check provenance. Even on puzzles. Especially on puzzles.

The Craftsman

England's answer to the showman was a quiet, self-taught puzzle writer named Henry Ernest Dudeney, and the contrast between the two men is the moral spine of this era. Where Loyd optimized for spectacle, Dudeney optimized for *elegance* — and he made elegance an explicit, almost mathematical criterion: the fewest pieces, the shortest path, the cleanest idea.

His masterpiece began as a challenge to his readers. On 6 April 1902, in his *Weekly Dispatch* column, Dudeney posed the haberdasher's problem: cut an equilateral triangle into as few pieces as possible that reassemble into a perfect square. That such a dissection exists was old news — a theorem of Wallace, Bolyai, and Gerwien (1807–1833) guarantees that *any* polygon can be cut into pieces that form *any* other polygon of equal area. Existence was free; the game was minimality. Four weeks later, on 4 May 1902, Dudeney published his answer: **four pieces**. Better: the four pieces can be *hinged* in a chain, so that the whole assembly swings one way into the triangle and the other way into the square, like a wooden animal turning inside out. In 1905 he showed a polished brass-hinged model to the Royal Society. It entered his classic collection *The Canterbury Puzzles* (1907), joined over the years by

the spider-and-the-fly (1903), *Amusements in Mathematics* (1917), and — in the *Strand*, July 1924 — the puzzle that became the archetype of an entire genre, SEND + MORE = MONEY. Several of these are lying in wait for you later in this book.

The haberdasher's problem also carries this chapter's longest fuse. Was four actually the minimum, or had Dudeney merely failed to find three? Nobody could say — for one hundred and twenty-two years. In December 2024, Erik Demaine, Tonan Kamata, and Ryuhei Uehara proved that no three-piece dissection exists (arXiv 2412.03865): Dudeney's four pieces are optimal, the first such optimality certificate ever obtained for a nontrivial dissection. Sit with that asymmetry. Finding the beautiful answer took one Edwardian columnist four weeks; proving that nothing better exists took mathematics twelve more decades. *Doing* and *proving you can't do better* are different sports, played on different timescales.

The two careers intertwined, not always with consent — Loyd reprinted Dudeney's inventions under his own name, one more entry in the provenance ledger. But the deeper legacy is Dudeney's standard: that a puzzle's answer can be not merely correct but *right* — minimal, surprising, and worth building in brass.

The Don at Play

While the showman and the craftsman fought for the newspaper public, an Oxford mathematics don was smuggling logic into the nursery. Charles Lutwidge Dodgson — Lewis Carroll to everyone since — treated the puzzle as a literary form. His Doublets, launched in *Vanity Fair* in 1879 (the idea, he said, came to him on Christmas Day 1877), asks the solver to turn one word into another, changing a single letter at a time, every intermediate rung a real word. It looks like a parlor game. It is training in making moves that are individually trivial but must remain globally legal — and, a century later, computer scientists would recognize word ladders as a textbook exercise in searching a graph. Carroll had invented an algorithms assignment before there was a discipline to assign it.

A *Tangled Tale* (1885) wrapped mathematics in fiction; *The Game of Logic* (1886) turned the syllogism into a children's board game. And in 1895, in the philosophy journal *Mind*, Carroll published the deepest four pages in this whole golden age: "What the Tortoise Said to Achilles." Zeno's two runners return — you met them a chapter ago — but this time the race is an argument. The Tortoise accepts every premise of a simple deduction but refuses to accept the *conclusion* until the rule licensing the inference is itself written down as a premise. Achilles obliges. The Tortoise then requires the rule licensing *that* — and so on, forever. The regress never ends, and its moral has never been improved on: the rules of inference cannot all be premises; justification must stop somewhere. Logic, the discipline that studies proof, cannot itself be proved all the way down. Working logicians still cite the children's author, without irony, on exactly this point.

Black Squares, Two Proofs

On 21 December 1913, Arthur Wynne, a journalist assembling the Christmas fun page of the *New York World*, published a diamond-shaped grid of interlocking words he called a "Word-Cross." The typesetter's transposition — *cross-word* — stuck. For a decade it remained a minor feature; then the craze of 1924–25 made it an industry, and the crossword settled into the position it has never surrendered: the first puzzle in history to become a *daily habit* for millions, puzzling as ritual rather than event.

Britain, characteristically, complicated it. From 1926, the setter Torquemada pioneered the cryptic crossword, where each clue is a tiny machine of misdirection; four decades later Ximenes codified its rules of fair play (1966). The cryptic's great invention is structural: every fair clue contains both a *definition* and an independent *wordplay* route to the same answer — two separate proofs, required to agree. The solver doesn't merely find an answer; the answer arrives pre-verified, redundancy built into the object itself. Engineers, accountants, and cryptographers had discovered the value of independent confirmation in their own trades; the cryptic crossword taught it to the breakfast table.

Into the Laboratory

Three quiet events, clustered near this age's end, changed what puzzles were *for*.

In 1901–02, in the *Annals of Mathematics*, a Harvard mathematician named Charles Bouton did something to the matchstick game of Nim that no one had ever done to any game: he *finished* it. A complete theory — from any position, who wins with best play, and how, no cunning or psychology required, the winning method a matter of holding a hidden balance invariant every turn. Strategy had become mathematics. The proof and the practical weapon wait in Chapter 9; what matters here is the precedent: a game could be *solved*, closed, known — which raised, faintly and for the first time, a question this book will meet again: what, then, was the human player contributing?

Between 1917 and 1945, psychology claimed the puzzle as an instrument. Wolfgang Köhler's apes stacked crates to reach hanging fruit (1917), and the moment of sudden reorganization he observed needed a name that wasn't "trial and error." Karl Duncker supplied the framework (in German in 1935; in English, posthumously, 1945): insight is the *restructuring of a representation* — the problem does not change; your description of it does — and its great enemy is *functional fixedness*, the inability to see past an object's usual role. The uses you know mask the uses you need. Duncker's most famous experiment — a candle, a box of tacks, a bare wall — appears as a puzzle in Chapter 4, so nothing more will be spoiled here. Max Wertheimer's *Productive Thinking* (1945) completed the movement. After three thousand years of sorting minds, puzzles were now *measuring* them.

And around 1945, a puzzle swept the wartime scientific world hard enough to become a security joke: given twelve coins, one of them counterfeit — heavier or lighter, unknown which — find it, and its direction, in three weighings on a balance. It devoured so many research-hours in Allied common rooms that colleagues joked about dropping it over Germany to paralyze the enemy's scientists. Freeman Dyson published the definitive analysis in the *Mathematical Gazette* in 1946. Its deep content was a new way of counting — not coins, but *possibilities against answers*: how much can three

verdicts of a balance possibly distinguish? That is information-theoretic thinking, practiced fluently two years before Claude Shannon gave information its mathematics and its name (1948). The full twelve-coin fight is a boss-room encounter in Chapter 7. Bring the counting.

The Age Closes

Run the arc once more. A French scholar gives puzzles provenance (1612); an encyclopedia gives them memory (1694); the penny press gives them the masses (1880); the masses are publicly defeated by an invariant they cannot see; a showman steals the credit and, more than a century on, a pair of historians take it back; a craftsman proves elegance is a standard, and mathematics spends twelve decades certifying him right; a don shows that logic itself can be played; the crossword makes puzzling a daily meal, with verification baked in; and in the last stretch, a game is solved outright, the "aha" becomes a laboratory measurement, and a coin puzzle teaches physicists to weigh information itself.

The puzzle entered this age as entertainment. It left as an instrument — for testing minds, measuring them, and asking what thinking is. The natural next step was audacious: stop studying the puzzles and start studying the solver. That project — thinking about thinking — is the next chapter, and the twentieth century pursued it with checklists, card-files, magazine columns, plastic cubes, and, at the very end, a riddle aimed at a new kind of traveler at the gate.

CHAPTER 3

Thinking About Thinking

For nearly four thousand years, puzzles tested minds. In the middle of the twentieth century, the enterprise turned around and began studying the thing it had been testing. The question of the new age was audacious and practical at once: what is solving? Is the knack a gift, or a craft with parts — nameable, teachable, trainable parts? Within a single decade, four very different answers appeared: a professor's checklist, an engineer's card-file, a columnist's stage, and — a little later — a logician's tavern. Between them they built the modern culture of thinking about thinking — the century in which "heuristic" left the seminar room, "lateral" acquired its modern meaning, and the puzzle column became a career — and this chapter follows that culture all the way to its strange present tense, where the oldest kind of riddle has returned to duty at the gate.

The Checklist

In 1945, the Hungarian-born mathematician George Pólya published a slim book with an immodest title: *How to Solve It*. Its claim was heresy against the romance of genius: solving is not lightning from nowhere. It has phases, and they can be written down. **Understand the problem. Devise a plan. Carry it out. Look back.** Under each phase, named tools — work backward from the goal; solve a simpler problem first; generalize; ask what is unknown, what is given, what connects them.

Written down like that, it looks almost too plain to matter. It mattered enormously. Pólya had done for problem-solving what Alcuin's title had only gestured at eleven centuries earlier: he made the *method* the explicit subject, separable from any particular problem and therefore teachable. The checklist is not an algorithm — nothing in it computes an answer — but it is a discipline aimed squarely at the two great amateur failure modes: diving in before understanding, and stopping the moment the answer appears.

Against the first, Pólya sets the humble interrogation every solver in this book performs on instinct: what is unknown? what is given? what connects them? Against the second, his most underrated instruction, the fourth: *look back*. The problem you have just solved is a specimen; mine it — for the trick that generalizes, for the check you skipped, for the shorter road you missed — or you have only won, not learned.

Every named-heuristics tradition since descends from that little volume, including the Moves running through this book. Chapter 8 takes Pólya into the arena and asks what champions had to add to make the checklist survive contact with a ticking clock — because knowing the phases, it turns out, is not the same as commanding them under pressure.

The Card-File

At almost the same moment, on the far side of the descending Iron Curtain, a Soviet patent engineer named Genrich Altshuller began reading. Starting in 1946, he worked through the patent archive — some two hundred thousand inventions, eventually — looking not at what each invention did, but at what *kind of trick* it pulled. His heresy was the mirror image of Pólya's: inventive insight is not even personal. Across unrelated industries, the same contradictions recur (stronger but lighter; faster but more precise), and they are resolved by the same few dozen maneuvers — segmentation, inversion, doing it in advance. Altshuller distilled roughly forty such inventive principles, arranged a matrix pairing recurring contradictions with the principles that dissolve them, and published the first paper on the method, with Rafael Shapiro, in 1956. The system became known as TRIZ, the Russian acronym for the theory of inventive problem-solving.

TRIZ's unsettling gift is captured in one sentence: *someone, somewhere, has already solved the abstract form of your contradiction*. Insight as lookup, not lightning. The romance drains out; a card-file takes its place; engineers mine it to this day. Between Pólya and Altshuller, the mid-century message was clear — the aha could be systematized. What it could not yet be was *popular*. That took a magician's touch.

The Stage

In December 1956, *Scientific American* ran a one-off article on hexaflexagons — folded paper strips that hide and reveal faces as they flex — by a Tulsa-born magazine journalist named Martin Gardner, a man with, famously, no mathematics degree at all. Readers wanted more. From January 1957, "Mathematical Games" ran monthly, and it ran for a quarter of a century, to 1981 — some three hundred columns that became, with no real competition, the greatest amplifier in the history of recreational thought.

The column worked because it was a bridge with traffic in both directions. Downward, the frontier came to breakfast: Gardner introduced the world to John Conway's Game of Life (October 1970), to Roger Penrose's aperiodic tiles (January 1977), and to RSA public-key cryptography (August 1977) — in a games column. Upward went the amateurs: readers' letters sharpened problems, found errors, occasionally made discoveries; a generation of working mathematicians would later date their vocation to a specific month of "Mathematical Games," and his readers still convene as the Gathering 4 Gardner, running since 1993. Gardner also performed an act of historical justice with direct bearing on the previous chapter: he revived Loyd and Dudeney for a new century — and reported their provenance honestly. His creed, stated many ways across twenty-five years, was that the line between play and mathematics is imaginary.

When he retired, Douglas Hofstadter's successor column bowed to him in its very name: "Metamagical Themas," an anagram of "Mathematical Games." It is hard to name another figure who created more solvers. The amateur renaissance of the late twentieth century — the math-circle kids, the puzzle collectors, the recreational-mathematics shelf at the bookstore — is, to a first approximation, Gardner's column metastasizing.

Two Prophets of Play

The popular age produced its own theorists, and honesty requires grading them differently.

In 1967, Edward de Bono, a Maltese-born physician, published *The Use of Lateral Thinking* and gave everyday language a term it clearly needed: the

sideways move, the deliberate escape from the groove of straight-ahead ("vertical") reasoning. His image was the hole: when you are digging in the wrong place, digging deeper does not help — the new hole must be started elsewhere. He built tools for forcing the sideways step — the provocation word PO (1972) and, most durably, *Six Thinking Hats* (1985), a meeting protocol in which a group wears one mode at a time (facts, feelings, caution, benefits, creativity, process) so that people stop defending identities and start taking turns at perspectives. The honest ledger, plainly stated: the empire was commercially vast, and the controlled evidence for a general, transferable "lateral thinking skill" stayed thin $\triangle$. The drills have real uses; the Six Hats survive as a genuinely clever de-ego-ing device; the brand oversold the science. Keep the exercises, grade the claims — a distinction this book will need again in Part V.

Raymond Smullyan needs no such asterisk. A logician who had worked as a stage magician, Smullyan published *What Is the Name of This Book?* in 1978 — a title that is itself a small self-referential trap — and turned a scrap of folklore — an island of knights who always tell the truth and knaves who always lie — into a systematic art form: self-reference, domesticated. The book is a ladder: each tavern riddle a rung slightly higher, until the reader who climbed for fun stands, without ever sitting through a lecture, at the edge of Gödel's incompleteness theorems — which *To Mock a Mockingbird* (1985) and *Forever Undecided* (1987) then walk right up to. Smullyan proved something lovely about the species: the deepest results in logic can be *walked to*, one puzzle at a time. The gymnasium had acquired a staircase to its roof. You will climb the lower rungs yourself in Chapter 5.

Between them, the two prophets mark the era's poles. De Bono preached breadth — escape the frame, whatever the frame. Smullyan practiced depth — follow one frame down to bedrock. The complete solver, this book will keep insisting, needs both hands.

The Object

In the spring of 1974, in Budapest, a young design lecturer named Ernő Rubik built a cube whose faces turned: fifty-four colored squares on six faces, each face free to rotate, the whole object somehow never falling apart.

Hungarian patent, 1975; world launch, 1980; and then the craze of 1980–82, the largest the genre had seen since the winter of 1880 — this time with the mathematics visible just beneath the plastic. The cube also changed the sociology of being stuck: for the first time, a puzzle's difficulty was carried in public. A scrambled cube on a desk is a legible fact about its owner, and for two years much of the planet walked around holding its unsolved state in its hands.

The cube has 43,252,003,274,489,856,000 positions. Against that number, two facts define the object's meaning. First: in 2010, Tomas Rokicki, Herbert Kociemba, Morley Davidson, and John Dethridge proved — with about thirty-five CPU-years of computation — that every one of those positions is solvable in at most **20 moves**. "God's number," the cubers call it. Second: no human method comes close. People solve the cube in fifty or sixty moves, by building small reusable tools — do-and-undo move sequences (the group theorist's *commutators*) that change a little while conserving the rest. The gap between 20 and 60 is the cube's real lesson, and it is one of the permanent lessons of this book: *knowing a solution exists is not the same as owning it*. Existence and findability are different currencies, and almost everything in Part III is about earning the second.

The Grid

Before the next relay race, notice a pattern the whole lineage obeys: every puzzle boom in this book rides a media technology. Papyrus carried Ahmes; court manuscripts carried Alcuin; cheap newsprint carried the 15 Puzzle and the crossword; the glossy monthly carried Gardner; molded plastic carried Rubik. The remaining booms of this chapter ride the web and the smartphone, and the pattern is worth keeping, because it predicts: whatever the next medium is, a puzzle will be among the first things it learns to do.

The same decades democratized the logic puzzle by way of an unlikely relay race. In 1979, Howard Garns, a retired American architect, published a modest number-placement puzzle called "Number Place" in a Dell magazine. It emigrated to Japan, where the publisher Nikoli refined it and, in 1984, named it — *sudoku*, "the digits must be single." It simmered there for twenty

years until Wayne Gould, a retired judge who had spent years writing a computer program to generate the puzzles, sold *The Times* of London on the idea; the first grid ran on 12 November 2004, and by 2005 the boom was global. The Lo Shu's grandchild — you met the ancestor two chapters ago — had become the daily meditation of hundreds of millions: constraint propagation as a breakfast ritual, no arithmetic required, every deduction forced.

The mathematics, characteristically, arrived after the masses. There are 6.67×10^{21} valid completed grids (Felgenhauer and Jarvis, 2005). The minimum number of clues that can force a unique solution is exactly 17 (McGuire, Tugemann, and Civario, 2012 — by exhaustive computer search). Computer science adopted the puzzle as a laboratory; Donald Knuth's "dancing links" made it a standard testbed for exact-cover algorithms. And a Tokyo mathematics teacher, Tetsuya Miyamoto, pushed the pedagogy one step further in 2004 with KenKen, built on his classroom philosophy — "the art of teaching without teaching": design the constraints well enough and the student instructs herself.

The Curriculum and the Room

Two smaller inventions of the era deserve their paragraphs, because each moved the gymnasium somewhere new.

In 2001, Colin Hughes launched Project Euler — the name saluting the master of Königsberg, whom you last saw in a one-sentence signpost in Chapter 1 — a numbered sequence of mathematical problems designed so that brute arithmetic fails and only *algorithmic insight* — the right representation, the right theorem, the right complexity class — runs in under a minute. That one-minute rule is a quiet piece of setter's genius: it draws a bright line between having an answer and having an *idea*, and it grades the difference automatically. Mathematics and programming, fused into a single self-paced curriculum. The puzzle had become a syllabus.

In 2007, in Kyoto, Takao Kato of the company SCRAP staged the first Real Escape Game — a room you are locked in, which is itself the puzzle — inspired by point-and-click browser games and built of physical props and

hidden mechanisms. Europe's first dedicated venue, Parapark, opened in Budapest (that city again) in 2011, and a global boom followed in 2013–15. What escape rooms added was everything print had to leave out: the body and the team. Search with your hands. Divide the roles. Reason under a clock, out loud, with friends. Anyone who has run one knows what the format really measures — not cleverness alone but the meta-skills around it: who searches, who synthesizes, who watches the time, who says out loud that the group has now spent nine minutes on the same locked cabinet. Print never tested that. Rooms do. Puzzling, solitary since Ahmes, had become architecture — and a social organism.

The Mirror

The lineage's newest mass ritual began, fittingly, as a love token. In October 2021, the software engineer Josh Wardle released a small web game he had built for his partner, Palak Shah: one five-letter word a day, six guesses, tiles turning green, yellow, or gray. Wordle's masterstroke arrived that December — a share button that posts your grid of colored squares *without the letters*: proof of work, spoiler-free. The grids flooded social feeds; by January 2022 the *New York Times* had bought the game for a price reported as "low seven figures."

Wordle's design lessons are a compressed course in why puzzles catch: scarcity (one a day — you cannot binge), community (everyone faces the *same* puzzle), and verification-without-spoilers (the grid shows *that* you solved, and how narrowly, never *what*). But its loveliest effect was epistemic. Within weeks, ordinary players were arguing about optimal opening words — which guess extracts the most information from the hidden answer? — and the arguments were, whether the arguers knew it or not, entropy calculations, made fully explicit in a widely watched 3Blue1Brown analysis (February 2022). A planet practicing information theory before breakfast. Dyson's wartime common room, scaled to a species.

The Gate Reopens

Which brings the lineage home.

In November 2019, the AI researcher François Chollet published a paper titled "On the Measure of Intelligence," and with it a test called the Abstraction and Reasoning Corpus — ARC. Its format is almost childishly humble: small grids of colored cells; a few examples showing a transformation, before and after; deduce the rule and apply it to a fresh case. Each task's rule is novel — invented so that no library of memorized patterns will contain it — and the rules lean only on the kind of knowledge a five-year-old owns: objects persist, shapes repeat, things fall and fill and mirror. Shown three examples, a person very often just sees it, and then struggles to say what exactly was seen. That gap — instant for the eye, awkward for the tongue — is the test's whole cunning.

Chollet's argument for this austerity cuts at the deepest assumption of the machine-intelligence age: *skill is not intelligence*. Any sufficiently trained system can be skilled at a task it has seen. Intelligence, he argued, is the *efficiency of acquiring new skill* — what you do when the rule is genuinely new. So the honest test is few-shot rule induction: tasks easy for humans and, by design, hard for machines.

Read that description again slowly. One question. Engineered so that a certain kind of mind slides straight off it. Posed at a gate everyone now wants to pass. Everything staked on the answer.

The lineage that opened at Thebes has closed its great circle: the gate-keeping riddle is back, and this time it is pointed at machines. Every gate asks three questions at once — who passes, who decides what passing means, and what passing actually proves — and the third has become, quite suddenly, one of the live intellectual questions of the age. Which minds pass, which slide off, how fast the scores are moving and what they mean — that is Part IV's story, told there with numbers and dates, because in that arena the numbers will not sit still. The Sphinx, at least, would find the whole arrangement familiar: a new kind of traveler on the road, an old kind of question in the way.

Before the scoreboard, though, the armory. Whatever species ends up reading it, the Moves are worth owning — and the next four chapters put them, one at a time, into your hands.

PART II

The Moves

CHAPTER 4

Breaking the Frame

Some puzzles resist you because they are hard. The puzzles in this chapter resist you because you are hard — because somewhere between reading the problem and attacking it, you quietly added a rule nobody stated, assigned an object a single job it never applied for, or drew a boundary that exists only in your head. The solution then sits in plain sight, on the far side of a wall you built yourself.

Psychology has names for this. Karl Duncker, working in the Gestalt tradition (German monograph 1935; English, posthumously, 1945), showed that objects get mentally welded to their usual function — he called it *functional fixedness* — and that the "aha" of insight is not a bolt of inspiration but a **re-structuring**: the same facts, suddenly held in a different representation. The Gestalt school had watched this in the raw. Köhler's chimpanzees (1917) sat stumped before out-of-reach bananas until, abruptly, a crate stopped being a thing-to-sit-on and became a thing-to-stand-on; Wertheimer's *Productive Thinking* (1945) argued that all genuine understanding has this shape — not adding facts, but re-seeing the ones you have. Puzzles became the laboratory instrument for measuring the phenomenon, which is why so many entries in this chapter carry an experiment in their provenance line.

The felt experience has a structure worth knowing in advance. First, fluency: the obvious approaches, tried quickly. Then the **impasse**: everything reasonable fails, and it feels like the problem is broken. This is the moment untrained solvers quit — and the moment the puzzle is actually asking its question, because an impasse on a solvable problem means exactly one thing: *some constraint in your head is not in the problem*. The re-structuring, when it comes, arrives all at once and feels less like computing an answer than like being handed one. That suddenness is why cultures file insight under inspiration, muses, and showers. The impasse is not failure. It is diagnostic — the exact sensation of a wrong frame refusing to contain a right answer.

So the muscle this chapter trains is not cleverness. It is *frame awareness*: the habit of noticing which constraints came from the problem and which came from you, and the willingness to put the second kind on the table and break them one by one. The puzzles are deliberately light — mostly one and two stars — because the difficulty here was never in the logic. It was in the seeing. Do not let anyone spoil these for you, and do not spoil them for yourself: on insight puzzles more than any other kind, the attempt is the training, and a peeked answer teaches nothing but trivia.

Two warm-ups

Both of these can be answered by a child and missed by a professor — and the professor's failure is the more instructive, because it shows that frames are not a knowledge deficit. More expertise means more defaults, installed deeper. That asymmetry is the lesson.

★ 11 • The Surgeon's Riddle

Statement. A father and son are in a car crash. The father dies; the son is rushed to hospital. The surgeon looks at the boy and says: "I can't operate on him — he's my son." How?

Pause here. The page can wait.

Solution. The surgeon is his **mother**. (Or, in the modern doubling: his other father.) Notice what just happened — or didn't. There is no logic step at all, no inference to perform. The "contradiction" was manufactured entirely by a default assumption — *surgeon = man* — that the reader's own schema inserted without asking permission. The riddle dates from the 1970s consciousness-raising era, and here is the uncomfortable part: it still works. A Boston University study around 2014 found most participants missing it, decades after the answer had entered folklore. Defaults do not expire on their own.

The Move — Audit the invisible assumption. The hardest constraints to drop are the ones you never wrote down. Before declaring a problem contradictory, list what you assumed *by default* — and interrogate the oldest defaults first, because those are the ones you can no longer see.

★ 12 • The Hotel and the Car

Statement. A man pushes his car to a hotel. When he arrives, he immediately knows he's bankrupt. Why?

Sit with it a moment — guess before you read.

Solution. He's playing **Monopoly**. His token is the car, he has landed on a hotel, and he cannot pay the rent. Every fact that seemed deranged — pushing a car, instant bankruptcy on arrival at a hotel — is perfectly mundane inside the right frame. That is the signature of this whole puzzle family (the "situation puzzle," canonized in Paul Sloane's lateral-thinking collections from 1991): the facts are individually bizarre but internally consistent, and no amount of patching the *current* context will reconcile them.

The Move — Switch the frame. When facts are strange but coherent, stop explaining them harder and start searching for the frame in which *all* of them are ordinary at once. Diagnosis, debugging, and detective fiction share this move: the symptom list that makes no sense in one story is trivial in another.

The classics of the invisible cage

The next four are the load-bearing classics of insight research and interview folklore alike. Each one plants a different species of false constraint: a boundary, a label, a dimension, a unit. Notice, as you work them, that the constraint is never stated — it is *suggested*, by the picture, the packaging, the tabletop, the word "measure." A well-built insight puzzle is a machine

for making you author your own cage, and the craft in the genre is entirely in that suggestion.

Duncker's experimental finding runs through all four: subjects who saw an object *in use* for its normal function solved less often than subjects who saw it idle (Adamson's 1952 replication confirmed the effect, and it has held up since). The frame arrives with the furniture. Which yields a practical corollary before you even meet the puzzles: when stuck, re-inventory the problem's raw materials as if they had just been dumped from a bag — unlabeled, unarranged, historyless.

★★ 13 • Nine Dots, Four Lines

Statement. Draw a 3×3 grid of nine dots. Connect all nine with **four straight lines** without lifting the pen.

Try it with a real pen. Paper is cheap; the jolt is not.

Solution. Watch your own first attempts: nearly everyone traces the perimeter and a diagonal, or zigzags row by row, and lands on five lines — always five, never four, always entirely inside the square. Inside the square that the dots suggest, it is impossible; you can convince yourself with a little case analysis. And that square is the whole trick, because the problem never mentions it. The eye supplies a boundary; the solver obeys a rule nobody stated. The canonical path leaves the grid twice: (1) start at the bottom-left dot and draw up through the left column, continuing one unit *past* the top; (2) come down diagonally through the top-middle and right-middle dots, continuing to one unit past the bottom-right; (3) run straight left along the bottom row; (4) go diagonally up through the center to the top-right dot. Four lines, nine dots, and two excursions into forbidden territory that was never actually forbidden.

The extensions are their own curriculum: with fat enough dots, three lines suffice; fold the paper and one line does it. Each further solution breaks another rule you didn't know you were obeying.

The Move — Question the imagined boundary. The constraint that blocks you may exist only in your representation. This puzzle — printed in Sam Loyd's *Cyclopedia* (1914) as the "Christopher Columbus egg puzzle," used experimentally by N. R. F. Maier in 1930, adopted by 1970s creativity consultants — is literally the origin of "thinking outside the box," a phrase it earned decades after its first printing. The cliché is tired; the move is not.

★★ 14 • Duncker's Candle Box

Statement. On a table: a candle, a book of matches, and a box of thumb-tacks. Attach the candle to the wall so that it burns without dripping wax on the table.

Stop. Picture the table. Give it three honest minutes.

Solution. Empty the tack box, **tack the box to the wall**, and stand the candle on it as a shelf. The false starts are predictable and instructive: tack the candle directly to the wall (it splits, and the tacks are too short); melt wax as glue (it won't hold the candle's weight). Both failures share one blindness — the box never registers as material at all. It is *packaging*, part of the problem's wrapping rather than the problem's inventory.

In Duncker's experiments this was measurable, and the measurement is the famous part: subjects who received the tacks *inside* the box solved the problem far less often than those who saw the box standing empty. Same objects, same task — the only difference was that in one condition the box was actively performing its default function. The box's role as container masked its role as platform. This is the most-cited experimental insight puzzle in psychology (Duncker, *On Problem-Solving*, 1935/1945), and it earned that citation count by being embarrassingly replicable: functional fixedness is not a quirk of careless people but a tax on everyone's perception. The fix costs one deliberate question, which is the Move.

The Move — Break functional fixedness. For every object in the problem, ask: what else is this, physically, ignoring its label? A box is also a shelf, a stand, a wall, a unit of measure. Labels describe history, not possibility.

★★ 15 • Six Matches, Four Triangles

Statement. With six matchsticks, form **four** equilateral triangles, each side one full match long. (Two triangles is easy; four seems to need nine matches.)

Reach for actual matches if you have them. Then pause.

Solution. Leave the tabletop: build a **tetrahedron** — three matches as a triangular base, three rising to meet at an apex. Four faces, six edges, exactly. The arithmetic of the impasse is worth spelling out: in the plane, each new triangle after the first two demands edges you don't have, and the count genuinely cannot close — flat, the puzzle is impossible. In three dimensions the shared edges come free. Nobody said the triangles had to lie down. You assumed the table because the matches were *on* one.

The Move — Add a dimension. When a problem is over-constrained, the constraint may be your ambient space, not your resources. The move has cousins everywhere: cutting a cake into eight pieces with three cuts, folding a proof, adding a time axis, introducing a new variable to break a deadlocked negotiation. Flatland is usually voluntary.

★★ 16 • Two Ropes, Forty-Five Minutes

Statement. Two ropes, each burning end-to-end in exactly 60 minutes — but *non-uniformly* (half a rope may take 5 minutes or 55). You have matches. Measure exactly **45 minutes**.

Pause here. The ropes aren't going anywhere.

Solution. The false start everyone makes: try to use rope *length* — burn to the halfway point, fold it to mark quarters, eyeball 45 as three-quarters. Non-uniformity kills all of it; the problem told you so in its own statement, and most solvers try anyway, because "measure" and "ruler" arrive welded together. Position on the rope tells you nothing about time. The insight is that one thing survives the mess: a rope lit at **both ends** finishes in exactly half its remaining burn time, wherever the flames meet, because the two burn-times must sum to the total and they run simultaneously. You cannot know *where* the flames will meet — and you do not need to. That's the restructuring: stop trying to defeat the non-uniformity and find the fact it cannot corrupt.

So: light rope A at both ends and rope B at one end, all at once. A's flames meet after exactly 30 minutes. At that instant, light B's other end: B has 30 minutes of burn left, and burning from both ends halves it — 15 more minutes. Total: $30 + 15 = 45$.

The Move — Exploit rates, not lengths. Non-uniformity destroys any measurement of *position*, but a flame from each end always splits remaining *time* in half. When a system is messy, hunt for the quantity the mess cannot touch — an invariant of rates. (Chapter 6 and Chapter 9 will make a whole arsenal out of invariants; this is your first taste of finding one under pressure.)

Words are frames too

Language is the frame you carry everywhere, which makes wordplay the cheapest insight-gym there is. A riddle is a sentence engineered so that its default reading is a trap — the verbal cousin of the nine-dot square. You met the form's terrifying ancestor at the gate of Chapter 1: the Sphinx's riddle worked by binding "legs" and "day" to their literal senses and daring you to re-map them. Nothing in the mechanism has changed in three thousand years; only the stakes have improved.

The two classics below train the same restructuring muscle as the matchsticks, except the material being restructured is grammar. They also pre-

view a discipline the deduction chapters will demand constantly: treating a sentence not as a vibe to be skimmed but as a structure to be *evaluated* — parsed, substituted into, reduced.

★★ 18 • Doublets: HEAD to TAIL

Statement. Turn HEAD into TAIL, changing one letter at a time, every intermediate step a real English word.

Pencil out a chain before reading on.

Solution. HEAD → HEAL → TEAL → TELL → TALL → TAIL. Five steps — Lewis Carroll's own chain. The strategy notes matter more than the answer: work from both ends toward the middle; the vowels are usually the hinge; and some words are dead ends simply because their one-letter neighborhoods are tiny. What you are actually doing — and Carroll could not have said it this way in 1879 — is graph search: states are words, edges are one-letter changes, and the puzzle asks for a path. Carroll published Doublets in *Vanity Fair* in 1879 (he claimed the idea came to him on Christmas Day 1877); computer-science departments now assign the identical exercise as "word ladders," a shortest-path homework problem. He invented an algorithms exercise a century early.

The Move — Walk legal neighbors. Progress by moves that are individually trivial but must remain globally legal at every step. This is the discipline of refactoring code that must keep compiling, of chess maneuvering, of any migration done live.

★★ 19 • "That Man's Father"

Statement. Looking at a portrait, a man says: "Brothers and sisters have I none, but that man's father is my father's son." Who is in the portrait?

Pause — and say your answer out loud before checking it.

Solution. His son. Most people answer "himself," and the error is precise: they collapse one level too early. Substitute from the inside out. "My father's son," given no siblings, is *me*. The sentence now reads: "that man's father is me." So the man in the portrait is his son. Natural language nests relations worse than algebra does — the words invite you to skim where a formula would force you to evaluate. This one circulated in nineteenth-century riddle books and became a Martin Gardner favorite precisely because it demonstrates, in fourteen words, how language obscures structure.

The Move — Substitute and simplify. Formalize nested relations and reduce them one binding at a time, exactly like evaluating an expression. Translate before trusting: if a sentence resists paraphrase, that resistance is where the trick lives.

The moves you now own

Eight puzzles, seven reusable moves — the first shelf of the toolkit this book is really selling:

- **Audit the invisible assumption** (11) — list your defaults; suspect the oldest ones most.
- **Switch the frame** (12) — when coherent facts stay strange, change stories, not effort.
- **Question the imagined boundary** (13) — ask which constraints are actually in the problem text.
- **Break functional fixedness** (14) — ask what each object is, not what it's *for*.
- **Add a dimension** (15) — the blocking constraint may be your ambient space.
- **Exploit rates, not lengths** (16) — find the quantity the mess cannot corrupt.
- **Walk legal neighbors / substitute and simplify** (18, 19) — small legal steps; evaluate language like an expression.

Notice that all seven are versions of one meta-move: *make the frame an object you can inspect*, instead of the lens you inspect everything else through. That is also why insight puzzles age so badly on second encounter — once seen, the frame can't be unseen — and why the training value lives in fresh impasses, not in memorized answers. When a real problem stalls, the drill is mechanical enough to run under stress: What am I assuming by default? What is each resource, apart from its label? Where exactly is the boundary of the legal space — who drew it? What would this look like one dimension up, or one frame over?

One honest caveat before you move on. Insight is the most romanticized of the thinking skills, and the research record is thinner than the aura — Duncker's restructuring effect is real and replicated; grand claims about trainable all-purpose "creativity" mostly are not (the full accounting, including what the brain-training industry got away with, waits in the Graveyard at the back of this book). What *does* transfer is the checklist you just built: notice the impasse, name the frame, and ask who installed it. The next chapter turns from breaking frames to filling them — deduction, where nothing is hidden except everything.

CHAPTER 5

Pure Deduction

Deduction puzzles hide nothing. Every fact you need is on the table from the first sentence; there is no trick frame to break, no pun to catch. The entire difficulty is discipline: holding a lattice of constraints in working memory, milking each fact completely before touching the next, and — above all — never confusing what you *know* with what you merely *assume*. Guessing before the forced moves are exhausted is the cardinal sin of this chapter, and nearly everyone commits it.

You have already done deduction's cousin once in this book: the wolf, the goat, and the cabbage back in Chapter 1 were state-space planning — mapping legal moves through a maze of constraints. This chapter turns from planning to inference. The material rises in a deliberate ladder: first liars and truth-tellers (the atom of case analysis), then wrongness as information, then the grand grid of the zebra puzzle, then puzzles where the *data* is *other people's ignorance*, and finally deduction against a clock.

One promise before we start. Each solution below teaches the path — including the wrong turns a smart solver actually takes — because in deduction, the method is the entire prize. The answers themselves are famously useless: nobody needs to know who owns a zebra.

Truth machines

Raymond Smullyan built a whole logical education (*What Is the Name of This Book?*, 1978) on one island: knights always tell the truth, knaves always lie. The setting sounds like a joke. It is actually a laboratory where every utterance becomes an equation.

★ 2 • Knights and Knaves: "We Are Both Knaves"

Statement. On an island, knights always tell the truth, knaves always lie. A says: "We are both knaves." What are A and B?

Pause here. The page can wait.

Solution. Run the only two cases. Suppose A is a knight: then his statement is true, so both are knaves — including A. Contradiction; knights don't lie. So A is a knave, and his statement is therefore false: they are *not* both knaves. Since A is one, B must be the knight. **A = knave, B = knight.**

Savor the fine grain, because it trips people for years afterward: a knave *can* say "we are both knaves" — precisely because the statement comes out false. He is lying about the pair, truthfully to his nature. The statement is not about what a knave would *never* say; it is about what must be false when he says it.

The Move — Assume, then break it. Hypothesize each case and let self-contradiction do the elimination. This is the atom of all logical deduction — proof by contradiction in its smallest wearable size.

★★ 4 • The Fork in the Road

Statement. A fork: one road to freedom, one to doom. Two guards, one always lies, one always tells the truth — you don't know which is which. One question, to one guard. What do you ask?

Stop and try to build the question yourself. It's buildable.

Solution. The false start: ask something factual ("Which road is freedom?") and you learn nothing, because you can't calibrate the answer without knowing the speaker. The insight is that you don't need to identify the liar — you need to *route your question through both guards at once*. Ask either guard: **"If I asked the other guard which road leads to freedom, which would he point to?"** — then take the *opposite* road. Check both cases: the truth-teller truthfully reports the liar's lie; the liar lies about the truth-

teller's truth. Either way the answer passes through exactly one lie, so both guards point wrong — reliably.

Smullyan's compact single-guard variant does the same with logical machinery: "Are you a knight *if and only if* this road leads to freedom?" Biconditionals also force the lie to cancel.

The Move — Fold the lie onto itself. You cannot remove an unknown distortion, but you can route the signal through it twice — or pair it against itself — so it cancels. Experimental science does the same thing and calls it differential measurement. (Provenance for the curious: folklore, with an early formal relative in Nelson Goodman's 1931 note; popularized by Smullyan and, for a generation of children, by the film *Labyrinth*, 1986.)

★★ 3 • The Three Mislabeled Boxes

Statement. Three boxes: one holds only apples, one only oranges, one a mix. They are labeled APPLES, ORANGES, MIXED — and **every label is wrong**. Drawing as few fruits as possible (no peeking), relabel all three correctly. How many draws?

One fruit? Two? Work it out before reading.

Solution. One draw suffices — but only from the right box: the one labeled **MIXED**. Its label is wrong, so that box is pure. Draw one fruit; say it's an apple. Then the box labeled MIXED is the apple box. Now the forced chain: the box labeled ORANGES cannot contain oranges (wrong label) and cannot contain apples (taken) — so it is the mixed box. The box labeled APPLES gets what remains: oranges. Symmetric if you drew an orange.

The instructive false start is drawing from the box labeled APPLES. Wrong label means it holds oranges *or* mix — one draw may not settle which, and the deduction stalls. The move that feels least informative — sampling the "mixed" box — is the most informative, because that label's wrongness is *total*: it pins the box to exactly two possibilities, and one fruit splits them.

The Move — Total wrongness is information. "All labels are wrong" is not chaos; it is a strong constraint. A guaranteed error is exactly as informative as a guaranteed truth. (A 20th-century folklore staple, beloved of Gardner-era collections and tech interviews, and secretly a lesson in where to spend your first measurement.)

The grand grid

Now the full orchestra. The logic grid asks you to hold five interlocking categories in your head at once, and it punishes every leak. What follows is the most famous specimen — solved completely, every step shown, because this walkthrough is the curriculum.

★★★ 5 • The Zebra Puzzle ("Einstein's Riddle")

Statement. Five houses in a row, each with a distinct color, nationality, drink, cigarette brand, and pet. The classic clue list (the 1962 original counted "there are five houses" as clue one, which is why it advertises fifteen):

1. The Englishman lives in the red house.
2. The Spaniard owns the dog.
3. Coffee is drunk in the green house.
4. The Ukrainian drinks tea.
5. The green house is immediately to the right of the ivory house.
6. The Old Gold smoker owns snails.
7. Kools are smoked in the yellow house.
8. Milk is drunk in the middle house.
9. The Norwegian lives in the first house.
10. The Chesterfields smoker lives next to the man with the fox.
11. Kools are smoked next to the house with the horse.
12. The Lucky Strike smoker drinks orange juice.

13. The Japanese smokes Parliaments.
 14. The Norwegian lives next to the blue house.

Who drinks water? Who owns the zebra?

This one deserves a table, a pencil, and half an hour. Take them.

Solution (unique — verified by exhaustive search). The path, with its two forks shown honestly:

Plant the anchors. Two clues are absolute: Norwegian in house 1 (clue 9), milk in house 3 (clue 8). Clue 14 then forces **blue = house 2** — the Norwegian's only neighbor. Three cells down, no guessing yet.

Squeeze the colors. Green sits immediately right of ivory (clue 5): the pair fits as (3,4) or (4,5), since house 2 is blue. Meanwhile red hosts the Englishman (clue 1), so red isn't house 1 (Norwegian) or house 2 (blue).

Here is the first fork, and the first lesson in *branch-killing*. Try ivory = 3, green = 4. Then red and yellow fill houses 1 and 5; the Englishman bars red from 1, so red = 5, yellow = 1. Kools = 1 (clue 7), horse = 2 (clue 11). Coffee = 4 (clue 3). House 1 smokes Kools, so it can't be the Lucky-Strike-and-orange-juice house (clue 12), and the Norwegian isn't the tea-drinking Ukrainian — so house 1 drinks **water**, and tea and orange juice land in 2 and 5. House 5 is the Englishman, so the Ukrainian and his tea sit in house 2, orange juice and Lucky Strike in house 5. The Spaniard and the Japanese fill houses 3 and 4; the remaining cigarettes — Old Gold, Chesterfields, Parliaments — fill houses 2, 3, 4. Now the collapse: the Old Gold smoker owns snails (clue 6), but house 2 owns the horse, and whichever of 3 or 4 isn't the Parliaments-smoking Japanese is the Spaniard — who owns the *dog*. Nowhere for the snails. The branch dies. Cross it out with satisfaction: a dead branch is a fact.

So ivory = 4, green = 5, and coffee = 5. Red and yellow fill 1 and 3; the Englishman again bars red from house 1, so **red = 3** (Englishman = 3), **yellow = 1**, Kools = 1, and the horse lives next door in house 2.

Propagate the drinks. House 1 smokes Kools, so not Lucky Strike, so no orange juice; the Norwegian is no Ukrainian, so no tea; milk and coffee are taken. **House 1 drinks water.** (First answer, and it arrived by pure elimina-

tion.) Tea and orange juice fill 2 and 4. A quick second fork: if the Ukrainian sat in house 4, orange juice would fall to house 2 along with Lucky Strike — leaving the Japanese, who smokes Parliaments, nowhere to sit among the remaining slots, and forcing the dog-owning Spaniard onto house 2's horse. Dead in two moves. So **Ukrainian = 2, tea = 2, orange juice = 4, Lucky Strike = 4**.

The last dominoes. Houses 4 and 5 take the Spaniard and the Japanese; Parliaments (clue 13) can't sit where Lucky Strike burns, so **Japanese = 5, Spaniard = 4, dog = 4**. Old Gold and Chesterfields remain for houses 2 and 3; snails follow Old Gold, and house 2 already keeps a horse — so **Old Gold = 3, snails = 3, Chesterfields = 2**. The fox lives next to the Chesterfields (clue 10): house 1 or 3; the snails hold house 3, so **fox = 1**. Four pets placed — the fifth cell is empty and waiting.

House	1	2	3	4	5
Color	yellow	blue	red	ivory	green
Nation	Norwegian	Ukrainian	Englishman	Spaniard	Japanese
Drink	water	tea	milk	orange juice	coffee
Smoke	Kools	Chesterfields	Old Gold	Lucky Strike	Parliaments
Pet	fox	horse	snails	dog	zebra

The Norwegian drinks the water; the Japanese owns the zebra — the two things the clues never mentioned, forced into existence by everything else.

The Move — Anchor, then propagate. Plant the absolute facts first, then squeeze every clue against the grid repeatedly; never branch while a forced move remains. And when you finally must branch, pick the cell with the fewest candidates and prosecute the branch to a quick death. This discipline — constraint propagation with lazy, reluctant branching — is exactly how modern SAT solvers earn their living.

Provenance, with a smile. This puzzle is universally sold as "Einstein's Riddle," sometimes with the garnish that "only 2% of people can solve it." Its actual birthplace: *Life International* magazine, December 17, 1962, no byline, solution and a list of successful solvers the following March. The grid mentions Kools, a cigarette brand of the 1930s, and ran seven years after Einstein died; the Lewis Carroll attribution drowns in the same bathtub. Einstein never saw it, 2% is marketing, and the puzzle needs neither — it is a masterpiece by an anonymous magazine staffer, which is somehow a better story.

Ignorance is data

The next rung looks impossible on first read: puzzles where people announce only what they *don't know* — and that turns out to be enough. The trick is to stop hearing "I don't know" as an apology and start hearing it as a mathematical statement about the speaker's information set.

★★★ 6 • Cheryl's Birthday

Statement. Cheryl gives Albert the month and Bernard the day of her birthday, chosen from: May 15, May 16, May 19, June 17, June 18, July 14, July 16, August 14, August 15, August 17. Albert: "I don't know when it is, but I know Bernard doesn't either." Bernard: "At first I didn't know, but now I do." Albert: "Then now I know too." When is Cheryl's birthday?

Resist the urge to skim. Each sentence is a filter — apply them in order.

Solution. The false start is trying to guess what Albert and Bernard are thinking. Don't. Translate each utterance into a fact about the candidate list, and cut.

Line 1, second half — "I know Bernard doesn't know." Bernard would know instantly only if his day were unique in the list: 19 (May only) or 18 (June only). Albert, holding just the month, is *certain* Bernard doesn't know — so Albert's month contains no unique day. May and June are eliminated wholesale. Left: July 14, July 16, August 14, August 15, August 17.

Line 2 — "At first I didn't know, but now I do." Bernard, holding the day, can now identify the date among the survivors — so his day appears exactly once in what remains. The 14 appears twice (July and August): gone. Left: July 16, August 15, August 17.

Line 3 — "Then now I know too." Albert, holding the month, is now certain — so his month contains exactly one survivor. August still holds two. **July 16.**

Notice what you just did: three sentences about *not knowing* pinned a single date out of ten, without anyone stating a fact about the birthday itself. Publicly declared ignorance is data. (And yes — the answer lands on a July 16, the very date on which this book's puzzle canon was compiled and verified. Coincidence, but the pleasing kind.)

The Move — Eliminate through others' knowledge. Each "I don't know / now I know" is a hard fact about the speaker's information set; apply it as a filter to the shared candidate list, in dialogue order. The same engine runs a five-star monster — Freudenthal's "impossible" sum-and-product puzzle — waiting in Chapter 7, alongside a hundred islanders who deduce their own eye color from silence itself.

Provenance. SASMO (Singapore & Asian Schools Math Olympiad), April 2015; it went globally viral via a Facebook post by TV host Kenneth Kong, briefly making epistemic logic the world's most argued-about topic at breakfast. The lineage runs straight back to Freudenthal's knowledge puzzles of 1969.

Deduction under a clock

Logic grids are static. The last rung adds scheduling — deduction about *sequences of actions* — where the trap is an assumption so reasonable it doesn't feel like an assumption at all.

★★★ 7 • The Bridge and the Torch

Statement. Four people must cross a rope bridge at night. The bridge holds at most two at a time; the group has one torch, and any crossing party must carry it. Crossing times: 1, 2, 5, and 10 minutes; a pair moves at the slower member's pace. Can everyone cross in **17 minutes**?

The obvious answer is 19. Pause until 19 bothers you.

Solution. The greedy plan — the 1-minute walker ferries everyone, since returns should be cheap — gives: 1&10 across (10), 1 back (11), 1&5 across (16), 1 back (17), 1&2 across (19). **19 minutes**, and most solvers stop here, certain the puzzle is misprinted.

The buried assumption is that the fastest person should always carry the torch. Question what the cost *structure* rewards instead: the 5 and the 10 are the expensive crossings, and in the greedy plan you pay for both separately. What if they crossed *together*, so the 10 absorbs the 5's cost entirely?

That requires staging: 1&2 across (2), 1 back (3), **5&10 across** (13), 2 back (15) — the 2 was parked on the far side precisely so a cheap courier would be available — 1&2 across (17). **Seventeen.** You spent one extra cheap return (the 2's) to merge the two worst costs into a single bill. Verified optimal — no strategy beats 17 (the full strategy space has been searched; this is Dijkstra's algorithm territory, if you enjoy overkill).

The Move — Make the expensive items share a bill. Interrogate the greedy assumption by asking what the cost structure actually rewards — here, co-scheduling the two worst costs so one hides inside the other. Batch your slow database queries; schedule both disruptive migrations for the same maintenance window. (Provenance: traceable to Levmore & Cook's *Super Strategies for Puzzles and Games*, 1981; a famous 1990s Microsoft interview question, back when interviews were fun.)

The moves you now own

Six puzzles, six moves — the rigor shelf of the toolkit:

- **Assume, then break it** (2) — case analysis with contradiction as the executioner.
- **Fold the lie onto itself** (4) — route a signal through its own distortion so it cancels.
- **Total wrongness is information** (3) — a guaranteed error is worth a guaranteed truth.
- **Anchor, then propagate** (5) — absolutes first; never branch while a forced move remains; kill branches fast.
- **Eliminate through others' knowledge** (6) — declared ignorance is a filter; apply in dialogue order.
- **Make the expensive items share a bill** (7) — audit what the cost structure rewards before trusting greed.

A closing observation that the next chapters will keep sharpening. Nothing in this chapter required brilliance — the zebra fell to bookkeeping, Cheryl to three filters, the bridge to one distrusted assumption. Deduction is the most learnable of the thinking arts precisely because its whole content is *procedure*: exhaust the forced moves, name your assumptions, treat ignorance as evidence. Machines, it should be said, do this flavor of reasoning superbly — grids and schedules are their home turf. The reason to train it anyway arrives whenever a problem hands you constraints without handing you a solver: a contract, a diagnosis, a timetable, a witness who only tells you what they don't know.

CHAPTER 6

Numbers, Chance, and Shape

The equipment is old. Human number sense counts four objects at a glance, feels the gap between twenty and forty but not the gap between twenty million and forty million, and files every probability under one of three headings: won't happen, might, will. That kit served foragers well. It still runs constantly — fast, fluent, and wrong in ways that carry no sensation of being wrong. The mistaken answer to every puzzle in this chapter arrives instantly, wearing total confidence.

The muscle trained here is the handoff: feeling the moment a question leaves intuition's jurisdiction and passing it to a stronger instrument — a notation, a computation, a proof. The instruments are not natural equipment. Positional arithmetic had to be smuggled into Europe inside a puzzle book (Chapter 1); probability theory is not yet four centuries old. Nobody perceives a conditional probability directly. The people who handle them correctly are the people who stopped trusting the feeling and learned to check.

Four movements: numbers, chance, estimation, shape — knowing exactly, knowing the odds, knowing roughly, and finally knowing that a thing is impossible. Along the way the gut will keep offering answers: 183 people, fifty-fifty, "the weekday obviously can't matter," "whichever treatment wins over-all is better." Each is wrong. Notice what the certainty felt like just before each correction; that feeling is the syllabus. Where gut and derivation disagree, the derivation wins. Every time.

Numbers

Arithmetic first: three puzzles about letting symbols carry what attention drops.

★ 26 • The Seventeen Camels

A father dies leaving seventeen camels: half to the eldest child, a third to the middle, a ninth to the youngest. Camels do not divide. While the heirs argue, a stranger rides up and lends them her camel. Eighteen camels: the eldest takes nine, the middle six, the youngest two — seventeen in all — and the stranger takes her own camel back and rides off. Every heir received at least the promised share. What happened?

Pause here. Where is the trick hiding?

Solution. Add the will's fractions: $1/2 + 1/3 + 1/9 = 9/18 + 6/18 + 2/18 = 17/18$. The will was never a partition of the estate. It under-allocates by exactly one eighteenth, and that gap is the whole mechanism. The borrowed camel completes the herd to a number the fractions bite on cleanly, and because the will only ever promised $17/18$ of a herd, the integer shares $9 + 6 + 2 = 17$ fit inside eighteen camels with one left over — the stranger's own. Each heir does better than the will's letter: nine camels against a promised 8.5, six against 5.67, two against 1.89. The trick is honest precisely because the original instruction was incoherent. The stranger didn't divide the estate; she repaired its arithmetic, and charged nothing. The tale is folklore across Arab, Indian, and European traditions — often told of the court wit Birbal — and Tartaglia was already discussing the seventeen-horse version in the sixteenth century.

The Move — Borrow a unit, return it. Temporarily extend a system to make its structure visible, then retract the extension. Mathematics does it constantly: add and subtract the same term, multiply by a cleverly written 1, adjoin a point at infinity, introduce a slack variable. The camel is a slack variable with legs.

★★★ 41 • SEND + MORE = MONEY

Each letter stands for a different digit; leading letters are not zero. Solve:

$$\text{SEND} + \text{MORE} = \text{MONEY}$$

Take a pencil. The carries are the whole game.

Solution. Name the carries c_1 (units into tens) up through c_4 (thousands into ten-thousands), and walk the columns from where the structure is stiffest.

MONEY has five digits, so the addition overflows: $c_4 = 1$, and since a carry never exceeds 1, **$M = 1$** .

Thousands column: $S + M + c_3 = O + 10 \cdot c_4$, so $S + 1 + c_3 = O + 10$. If $c_3 = 0$, then $S = O + 9$, forcing $O = 0$, $S = 9$. If $c_3 = 1$, then $S = O + 8$, so O is 0 or 1 — and 1 is taken. Either way **$O = 0$** , with $S = 9$ or 8 pending.

Hundreds column: $E + O + c_2 = N + 10 \cdot c_3$, which is $E + c_2 = N + 10 \cdot c_3$. If $c_3 = 1$, the left side must reach $10 + N$; the only escape is $E = 9$, $c_2 = 1$, $N = 0$ — and 0 belongs to O . Dead. So **$c_3 = 0$** , which settles **$S = 9$** , and leaves $N = E + c_2$. Letters denote different digits, so $N \neq E$, forcing $c_2 = 1$ and **$N = E + 1$** .

Tens column: $N + R + c_1 = E + 10 \cdot c_2 = E + 10$. Substitute $N = E + 1$ and it collapses to $R + c_1 = 9$. $R = 9$ is taken, so **$R = 8$** and $c_1 = 1$.

Units column: $D + E = Y + 10$. The digits left are $\{2, 3, 4, 5, 6, 7\}$; E and $N = E + 1$ must be consecutive picks from them, and $D + E$ must reach at least 12. Test the candidates: $E = 5$ ($N = 6$) demands $D = Y + 5$, and from the remaining $\{2, 3, 4, 7\}$ that gives $D = 7$, $Y = 2$ — everything fits. Every other choice of E strands D above the digits that remain. The unique solution, confirmed by exhaustive search: **$9567 + 1085 = 10652$** . Dudeney published the puzzle in the *Strand Magazine*, July 1924; the genre got its name — the cryptarithm — in 1931.

The Move — Squeeze the tightest constraint. Don't start where filling is easy; start where the structure is most rigid — here, the overflow into a fifth digit — and let the forcing cascade. In any constraint problem, most-constrained-variable-first is the professional's default. Amateurs start in the friendly corner.

★★★ 48 • The Self-Describing Number

Find the 10-digit number whose first digit says how many 0s the number contains, whose second digit says how many 1s, and so on, the tenth digit counting the 9s.

Build it digit by digit. The number polices itself.

Solution. Call the digits d_0 through d_9 , where d_i counts the occurrences of digit i . Two facts discipline the search. First, the digits sum to 10: the census counts each of the ten positions exactly once. Second — counting another way — the sum of the number's digit values equals $\sum i \cdot d_i$, and that is the same sum again: 10. Two totals, both 10, so the number is mostly zeros: a large d_0 and a light tail.

Try $d_0 = 7$. Then digit 7 appears, so $d_7 = 1$, and the tail beyond d_0 must sum to 3. That 1 sitting at position 7 means the number contains a 1, so $d_1 \geq 1$. But $d_1 = 1$ is self-defeating — writing it creates a second 1 — and $d_1 = 2$ summons a 2 into existence, needing $d_2 = 1$, and the total hits $7 + 2 + 1 + 1 = 11$. Dead.

Try $d_0 = 6$. Then $d_6 = 1$, tail sum 4. The 1 at position 6 forces $d_1 \geq 1$; set $d_1 = 2$, which puts a 2 in the number, so $d_2 = 1$ — and the census closes: $6 + 2 + 1 + 1 = 10$. Read it off: **6210001000**. Check it against itself: six 0s, two 1s, one 2, one 6, nothing else. It holds, and machine search confirms it is the only such number in base 10. The family is called the autobiographical numbers — recreational folklore of the Gardner era.

The Move — Seek the fixed point. When the description is part of the described, demand $x = f(x)$ and let consistency do the eliminating. Fixed points are where self-reference stops being a parlor trick and becomes mathematics: Gödel numbering, quines, Nash equilibria, and every "this sentence..." that actually works.

Chance

Probability is where confident people go to be wrong. The gut offers three settings — won't, might, will — and then backfills a story. The mathematics

that corrects it is barely 370 years old, and it has never become instinct in anyone. For the four puzzles that follow, do one extra thing: before solving, say the instant answer out loud and write it down. The point of this section is not the right answers. It is the documented experience of the wrong ones.

★★ 32 • The Birthday Paradox

How many people must be in a room before it is more likely than not that two of them share a birthday?

Commit to a number before reading on.

Solution. The instinct says something near 183 — half of 365. The answer is **23**. Compute the complement, the probability that all birthdays differ: the second person avoids the first with probability $364/365$, the third avoids both with $363/365$, and so on. For 23 people, $1 - (365/365)(364/365)\cdots(343/365) = \mathbf{0.5073}$; at 22 people it is 0.476, so 23 is the crossing. Intuition breaks because it counts *your* potential matches — 22 chances, which feels rightly small. But the event lives on **pairs**, and 23 people carry $C(23,2) = 253$ of them, each a lottery ticket nobody is watching. The growth is brutal: by 50 people the probability is 97 percent; by 70, it is 99.9. The result is usually credited to Richard von Mises (1939), carried into folklore by Feller's probability textbook and Gardner's columns.

The Move — Count the complement, and count pairs. "At least one" almost always means: compute *none*, subtract from 1. And when items interact, the scale of the problem is n^2 , not n — the arithmetic behind hash collisions, DNA-database coincidences, and why "small world" stories are cheap.

★★★ 33 • Monty Hall

Three doors. Behind one, a car; behind the other two, goats. You choose a door. The host — **who knows where the car is, must open one of the doors**

you didn't pick, and must show a goat — opens one: goat. He offers you a switch to the last closed door. Do you take it?

Decide now — switch or stay — and rate your confidence.

Solution. Staying wins $1/3$ of the time; switching wins $2/3$. Nearly everyone's gut says two closed doors, fifty-fifty — the mailbag below proves how strong the pull is. Enumerate, having picked door 1. Car at 1 (probability $1/3$): the host opens 2 or 3; switching loses. Car at 2 ($1/3$): he is forced to open 3; switching wins. Car at 3 ($1/3$): forced to open 2; switching wins. Two worlds out of three, the switch takes the car.

The fifty-fifty feeling treats the host's reveal as scenery. It is not. Your first pick is wrong $2/3$ of the time, and in every one of those worlds the host's hand is forced — his choice concentrates the entire remaining $2/3$ onto the door he was not allowed to open. Amplify until the signal is deafening: a hundred doors, you pick one, and the host, knowing, swings open ninety-eight goats and leaves one door shut. Your door keeps its $1/100$. The survivor carries $99/100$. Nobody feels fifty-fifty at a hundred doors; three doors is the same machine with the volume turned down.

The constraint is everything. Contrast "Monty Fall": a host who slips, opens one of the other doors *at random*, and happens to reveal a goat. The stage looks identical — and now the answer genuinely is $1/2$. Same picture, different procedure, different probability.

History, for consolation. Steve Selvin posed the problem in *The American Statistician*, 1975 (its grandfather is Bertrand's box paradox, 1889). When Marilyn vos Savant published the correct answer in *Parade* in September 1990, roughly ten thousand letters arrived telling her she was wrong — around a thousand from PhDs. She wasn't. Paul Erdős reportedly stayed unconvinced until shown a computer simulation. If the puzzle beat Erdős's intuition, it was allowed to beat yours. What matters is which instrument gets trusted next time.

The Move — Condition on the procedure, not the picture. Evidence is not what is seen; it is how likely that sight was under each hypothesis, given the process that produced it. Identical observations from different protocols carry different information. Ask what machine generated the reveal — in boardrooms and biopsy results as much as game shows. This may be the most practically valuable probability habit in the book.

★★★ 35 • Simpson's Paradox

Two treatments for kidney stones, real clinical data. Open surgery beats keyhole surgery on small stones *and* on large stones. Overall, keyhole wins. All six numbers are correct:

Success rate	Small stones	Large stones	All patients
Open surgery	93% (81/87)	73% (192/263)	78% (273/350)
Keyhole (PCNL)	87% (234/270)	69% (55/80)	83% (289/350)

Which treatment is better — and how can the numbers even do this?

Check the arithmetic yourself. It all holds.

Solution. Nothing is wrong with the arithmetic; something is wrong with the question "which row wins?" asked of the last column. The reversal comes from the **case mix**. Open surgery was assigned mostly the hard cases — 263 of its 350 patients had large stones — while keyhole got mostly easy ones, 270 of 350 small. The "All patients" column therefore compares open-surgery-on-hard-cases against keyhole-on-easy-cases, and severity, not treatment, drives the totals. Severity is a confounder: it steered the choice of treatment and it worsens the outcome. Given that, the stratified columns are the honest comparison — open surgery wins — and the aggregate answers a different question, usually the wrong one. The data are from Charig et al., BMJ, 1986; the effect carries Simpson's name (1951) with Yule's priority (1903), and it walks the earth: the 1973 Berkeley admissions case, where

women applied to the more competitive departments (Bickel et al., *Science*, 1975), batting averages, and every A/B test segmented after the fact.

The Move — Stratify before comparing. Check the denominators; hunt for the third variable steering cases into groups. And respect the converse discipline: aggregation is never neutral — *someone chooses the level at which the truth gets read*. When a summary and its parts disagree, ask which one answers the question actually being asked.

★★★★ 34 • The Tuesday Boy

A parent says: "I have two children. One of them is a boy born on a Tuesday." What is the probability both children are boys? And — seriously — does the Tuesday matter?

Your first answer is wrong. Probably your second, too.

Solution. The gut says the weekday is decoration: one is a boy, so the other is a boy with probability $1/2$. A better-trained gut recalls the classic two-children puzzle and says $1/3$. The enumerated answer is $13/27 \approx 0.481$ — and the Tuesday is doing real work.

Take the sample space seriously: each child is a sex-and-weekday pair, 14 equally likely kinds, so an ordered pair of children comes in 196 kinds. Families with at least one Tuesday-boy: $14 + 14 - 1 = 27$ (first child a Tuesday-boy, or second, minus the one family counted twice — where both are). Among those 27, the both-boy families number $7 + 7 - 1 = 13$. Hence $13/27$: strictly between $1/3$ and $1/2$. The mechanism is that subtracted 1. The more specific the mentioned detail, the closer the answer creeps to $1/2$, because a specific enough detail effectively *names which child* is being discussed — and once one child is pinned, the other is a plain coin flip. "One is a boy" leaves maximal ambiguity ($1/3$); "one is a boy born at 6:52 on a Tuesday morning" is nearly a proper name ($\approx 1/2$); Tuesday alone sits between.

Now the required honesty. $13/27$ answers a *census* question: among all two-child families having at least one Tuesday-born boy, what fraction have two

boys? If instead a parent simply volunteers a fact about one child, the selection procedure is different and so is the number — typically straight back to $1/2$. The puzzle has no unique answer until the sampling procedure is stated $\triangle$. Gardner, who introduced the underlying boy-girl paradox in 1959, later retracted his own $1/3$ as ambiguous for exactly this reason. Gary Foshee detonated the Tuesday version from the stage of the Gathering 4 Gardner in 2010, and the worldwide argument it started was, itself, the lesson.

The Move — The sampling defines the odds. Before computing any conditional probability, write down the procedure that produced the statement. Data never speaks alone; it speaks through its selection mechanism — the same lesson that returns as survivorship bias, Berkson's paradox, and half of all misread statistics.

Estimation

Past the edge of available data lies most of real life. What replaces exactness there is not guessing; it is controlled roughness.

★★ 46 • Fermi's Piano Tuners

How many piano tuners work in Chicago? No looking anything up — that restriction is the whole exercise.

One page, a pencil, no search bar.

Solution. Build a chain of factors, each defensible to within a factor of two or three. Chicago: about 3 million people. At roughly 3 per household, call it 1 million households. Perhaps 1 in 20 keeps a piano: 50,000 pianos, with schools, churches, and concert halls adding a few thousand more. A piano wants tuning about once a year, so $\sim 50,000$ tunings a year. One tuner manages maybe 4 a day across 250 working days: 1,000 a year. Demand over capacity: $50,000 \div 1,000 \approx 50$ **tuners**. The honest answer is a band, not a point — 30 to 150, order 10^2 — and the old Chicago phone directories

agreed, listing dozens to a hundred-odd. Enrico Fermi ran his Chicago classroom on questions like this in the 1940s; the same reflex let him estimate the Trinity bomb's yield on 16 July 1945 from how far the blast wave pushed scraps of falling paper.

Why it works: each factor is soft by a factor of ~ 2 , but the errors are independent — some high, some low — and they partially cancel in the product. The order of magnitude survives even though every input is a guess. And the discipline pays twice: an estimate made *before* a serious calculation is the alarm that later catches the calculation's slipped decimal point.

The Move — Decompose, estimate, multiply. Convert "unknowable" into a chain of individually guessable factors; think in powers of ten; state the band and refuse the false point. Confidently within threefold beats precisely wrong by a million.

Shape

The eye is a confident witness and a poor one. Two closers: a puzzle where the picture lies, and one where a two-line argument outsees every picture that could ever be drawn.

★★★ 22 • The Missing Square

Four flat pieces: a right triangle 8 units long and 3 high; a right triangle 5 long and 2 high; and two L-shaped pieces of 7 and 8 unit cells. Arranged one way, they fill what appears to be a right triangle 13 units long and 5 high. Rearranged, the same four pieces form the same 13-by-5 "triangle" — with one interior unit cell empty. Cut them from graph paper: both assemblies look impeccable. Where did the square go?

Where can a whole square hide?

Solution. In the hypotenuse. Neither figure is a triangle. The 8×3 piece's slanted edge rises at slope $3/8 = 0.375$; the 5×2 piece's at $2/5 = 0.4$. Laid end to end they form a line that *bends* — by barely more than a degree, well un-

der what the eye resolves at pencil width. In one arrangement the false hypotenuse sags concave, and the figure holds area 32; in the other it bulges convex, area 33. A true 13×5 triangle would hold $13 \times 5 \div 2 = 32.5$: the two impostors sit half a unit on either side of it. The missing square never left. It is smeared along the seam as a sliver-thin parallelogram of area exactly 1. Paul Curry, a New York amateur magician, built the effect in 1953; Gardner gave it the classic exposition in *Mathematics, Magic and Mystery* (1956); its grandfather is Sam Loyd's vanishing-warrior card *Get Off the Earth* (1896).

The Move — Verify the diagram. A picture is a claim, not a proof. Sum the areas; check the slopes; measure. The magician's version plays fair — it announces itself as a trick. The chart in the slide deck that shows exactly what everyone hoped to see makes no such announcement.

★★ 20 • The Mutilated Chessboard

Take a chessboard and cut away two diagonally opposite corner squares, leaving 62. You have 31 dominoes, each covering exactly two edge-adjacent squares. Can the 31 dominoes tile the mutilated board?

Try a few tilings in your head before answering.

Solution. No — and the proof never tries a single tiling. The two removed corners share a color; a chessboard's opposite corners always do. What remains is 32 squares of one color and 30 of the other. Every domino, wherever it lies, covers exactly one dark and one light square — adjacency guarantees it. So 31 dominoes cover 31 dark and 31 light. But the board offers 32 and 30. Done.

Feel what just happened. By hand, attempt after attempt would fail without ever revealing whether the failure was the board's or the attempt's. The coloring argument refutes *all* arrangements at once — it examines zero of them — by exhibiting a quantity that no legal placement can change. This is why John McCarthy proposed the puzzle as a benchmark for automated theorem-provers: a machine that only searches arrangements cannot see a coloring. The puzzle was posed by the philosopher Max Black (*Critical*

Thinking, 1946) and spread by Gamow & Stern and by Gardner's 1957 column, and it remains the cleanest two-sentence aha in mathematics.

The Move — Find the invariant. To prove a thing impossible, find a quantity every legal move conserves and show that start and goal disagree on it. Chapter 9 forges this single flash into an arsenal — colorings, parities, and conserved sums that execute whole families of impossible tasks in a paragraph.

The moves you now own

- **Borrow a unit, return it** (★ 26, the camels) — extend the system, read its structure, retract the loan.
- **Squeeze the tightest constraint** (★★★ 41, SEND+MORE) — start where the structure is rigid and let the forcing cascade.
- **Seek the fixed point** (★★★ 48, the self-describing number) — when the description describes itself, demand consistency: $x = f(x)$.
- **Count the complement, and count pairs** (★★ 32, birthdays) — "at least one" means compute "none"; interactions scale as n^2 .
- **Condition on the procedure, not the picture** (★★★ 33, Monty Hall) — evidence is how likely the sight was under each hypothesis.
- **Stratify before comparing** (★★★ 35, Simpson's) — find the variable steering cases into groups; distrust the flattering aggregate.
- **The sampling defines the odds** (★★★★ 34, the Tuesday boy) — no conditional probability without the generating procedure.
- **Decompose, estimate, multiply** (★★ 46, piano tuners) — powers of ten; bands, not points.
- **Verify the diagram** (★★★ 22, the missing square) — a picture is a claim.
- **Find the invariant** (★★ 20, the chessboard) — impossibility by conservation.

One meta-move rode along with all ten. Every failure of intuition in this chapter announced itself as certainty — the answer that felt most obvious

was the trap, every time. The stronger the click of obviousness around a number, the more it has earned the two-line check.

CHAPTER 7

The Boss Room

These are the puzzles mathematicians tell each other. Each has stopped a common room full of professionals; several have grown academic literatures; the last has resisted every proposed resolution for nearly eighty years and closes the chapter still open. The ratings here read ★★★★★ and ★★★★★, and they measure depth, not prerequisites — nothing in this room needs more than arithmetic. What each puzzle demands is a move that is probably not in the kit yet, executed down a long chain with zero leaks.

The room has rules. One puzzle per sitting. Read a statement, then close the book and carry the problem for a day before reading on: the value is manufactured during the attempt, not at the answer. Solving one of these cold is rare, and it is not the bar. The bar is meeting each solution with enough failed approaches behind you that the key idea, when it arrives, answers a question you had actually asked. A boss studied transfers nearly as well as a boss beaten.

One observation before the door opens. Nearly every scenario here threatens its characters — prisoners, pirates, islanders bound by a decree, a condemned man. The genre is being honest about the altitude: at this level there is no partial credit. Six of the seven solutions below are airtight, certainty by construction. The seventh trades certainty for a 31 percent miracle — and then proves that no strategy on earth does meaningfully better.

★★★★ 8 • The Twelve Balls

Twelve visually identical balls; exactly one is off-weight — heavier or lighter, unknown which. A balance scale, and three weighings only. Identify the odd ball *and* whether it is heavy or light, guaranteed.

This one repays a full evening. Give it one.

Solution. Before any cleverness, budget. There are 24 possible answers: twelve balls, each possibly heavy or light. A weighing has exactly three outcomes — left heavy, right heavy, balance — so three weighings distinguish at most $3^3 = 27$ futures, while two manage only 9. Three is the floor, and $24 \leq 27$ says the floor is *barely* enough: no weighing may waste information. After every outcome the surviving cases must number at most 9, then 3, then 1.

That arithmetic designs the first move by itself. Weigh six against six and balance is impossible: a tilt is certain, twelve cases survive it, and twelve cannot fit through the 9 futures two weighings offer. Dead before starting. Five against five? A tilt leaves ten. Dead. Four against four: a tilt leaves eight (four suspects of heaviness, four of lightness), and balance leaves eight (four untouched balls, either direction). Alive everywhere. The information bound forced the 4–4 split before ingenuity was even consulted.

First weighing: {1,2,3,4} against {5,6,7,8}.

If it balances — the odd ball is among 9–12, eight cases. Second weighing: {9,10,11} against {1,2,3}, three suspects against known-good ballast. Balance → 12 is odd; weigh 12 against any good ball to read its direction. Left heavy → one of 9, 10, 11 is heavy; weigh 9 against 10 — the heavier pan names it, balance convicts 11. Right heavy → the same three are light suspects; mirror.

If the left side sinks — either 1–4 is heavy or 5–8 is light, eight cases. The second weighing must mix suspects across the pans: **{1,2,5} against {3,4,6}**, parking 7 and 8. Balance → 7 or 8 is light; weigh 7 against 8, and the lighter is guilty (they cannot balance). Left sinks again → 1 or 2 heavy, or 6 light; weigh 1 against 2 — heavier is guilty, balance convicts 6 light. Right sinks → 3 or 4 heavy, or 5 light; weigh 3 against 4 — heavier is guilty, balance convicts 5 light. Count the branches: $2 + 3 + 3 = 8$. Every case lands.

If the right side sinks — the mirror image.

Two ideas carried the design. Cleared balls become instruments — known-good ballast that lets three suspects be weighed against "truth." And suspects rotate between pans so that each outcome splits the survivors nearly in three. This is the problem wartime mathematicians joked about air-drop-

ping over Germany to paralyze enemy research; Freeman Dyson's systematic analysis ran in the *Mathematical Gazette*, 1946 — information-theoretic thinking two years before Shannon gave it a name.

The Move — Maximize information per test. Budget outcomes before designing experiments, then shape every test so its outcomes are as equally loaded as possible. This is experiment design in miniature, and it reappears wherever questions cost money: debugging, medical triage, twenty questions of every kind.

★★★★ 44 • The Pirate Game

Five perfectly rational pirates, strictly ranked: A most senior, then B, C, D, E. They must divide 100 gold coins. The most senior pirate proposes a split; all five vote; a tie passes, the proposer's vote carrying. If the proposal fails, the proposer is thrown overboard and the next in rank proposes. Every pirate's priorities, in order: survive; then maximize gold; then, all else equal, *enjoy* seeing a rival thrown overboard. What does A propose?

Take your time. The gold isn't going anywhere.

Solution. Forward reasoning drowns — "what will B do, which depends on what C would do, which depends..." spirals without anchor. Anchor at the end, where the game is trivial, and roll back.

Two pirates left (D, E): D proposes 100–0 for himself, votes yes, and the tie passes. E's opinion is decorative.

Three (C, D, E): if C dies, D's reign gives E nothing — so E's vote costs exactly one coin, which beats D's promised nothing (the bloodlust tiebreak only applies *all else equal*). C proposes **99–0–1** and it passes 2–1. D cannot counterbid: his own next-round future is 100, so C never buys him.

Four (B, C, D, E): if B dies, C's regime pays D zero. D is now the cheap vote: **99–0–1–0** passes 2–2 on B's casting vote. Buying E instead would cost 2, since E already collects 1 under C.

Five: if A dies, B's plan pays C zero and E zero. A needs three votes of five — his own plus two others — and the two paupers-under-B are exactly C and E, one coin each. **A proposes 98-0-1-0-1**, and it passes 3-2.

The shock is the direction. The most exposed player — first to face the plank — captures 98 percent, while the pirates with "obvious" leverage collect one coin or none. Voting rules plus common knowledge of rationality allocate power in ways no intuition about seniority predicts. A second lesson comes cheaper here than in a live negotiation: the theorem is fragile. Real humans reject insulting offers out of spite — ultimatum-game experiments show it reliably — so backward induction holds exactly as far as its premises do. Know when the game is pirates and when it is people. The problem is economics folklore; the canonical telling is Ian Stewart's "A Puzzle for Pirates," *Scientific American*, May 1999, crediting Stephen Omonero.

The Move — Solve the endgame first. Anchor at the final subgame, where the answer is trivial, and roll toward the present, pricing every vote at its owner's alternative future. Chess endings, contract design, and any negotiation with a walk-away run on this engine.

★★★★ 9 • 100 Prisoners and a Light Bulb

One hundred prisoners. One interrogation room holding a single light bulb, initially off. Each day the warden brings one prisoner, chosen uniformly at random, to the room; the prisoner may toggle the bulb or leave it alone. At any moment, any prisoner may declare: "All one hundred of us have visited this room." If the claim is right, everyone goes free; if wrong, everyone dies. The prisoners may confer once, before the first visit. Find a strategy that cannot be wrong.

Design your own protocol before reading theirs.

Solution. The room's shared memory is one bit. Nobody can count in it — but a hundred people can *talk through* it, if they fix in advance who says what, and when. Appoint one prisoner the **Counter**. Everyone else follows a

single rule: the first time you enter and find the bulb off, switch it on — once in your life; in every other circumstance, touch nothing. The Counter's rule: whenever he finds the bulb on, he switches it off and adds one to a private tally. When the tally reaches 99, he declares.

Correctness is absolute, not probable. Each on-to-off cycle the Counter records corresponds to exactly one distinct non-counter's first-and-only signal: nobody signals twice, and no signal can be lost, because only the Counter ever switches the bulb off. Tally 99 therefore means 99 distinct others have visited — and the Counter himself makes 100. The bulb is a one-bit channel; the protocol serializes 99 messages through it, one at a time.

The price is patience. Simulation puts the expected wait near 10,300 days — twenty-eight years — most of it spent with the bulb burning idle, waiting for the Counter's next random visit. Faster protocols exist: multi-stage schemes in which prisoners carry "tokens" of escalating denomination cut the expected time substantially (the family is analyzed by Dehay, Ford & Segerman, *Mathematical Intelligencer*, 2003). But the single-counter scheme is the canonical proof of possibility, and its correctness argument fits in four sentences — its own kind of optimal. The puzzle is young for a classic: first posted to the newsgroup rec.puzzles on 24 July 2002.

The Move — Design a one-bit protocol. When shared state is tiny, stop asking "what can I deduce?" and ask "what protocol can we agree on beforehand?" Designate an aggregator; make every signal single-shot and unambiguous. This is distributed-systems engineering told as a fable.

★★★★★ 37 • 100 Prisoners and 100 Boxes

The same hundred prisoners, a crueler room. One hundred boxes in a row; inside them, the numbers 1 to 100 in a uniformly random permutation, one per box. Each prisoner, alone, may open fifty boxes in search of his own number. The room is then restored exactly as it was, and he leaves by another door, able to signal nothing. If *every* prisoner finds his own number, all go free; a single failure kills them all. Opening boxes at random, each

prisoner succeeds with probability $1/2$, so the group survives with probability $(1/2)^{100} \approx 8 \times 10^{-31}$ — call it never. There exists a strategy whose survival probability exceeds 31 percent. Find it.

Read that again: from 10^{-30} to nearly one in three. Now hunt.

Solution. No communication is possible, so the coordination must be smuggled through the only shared structure that exists: the contents of the boxes themselves. **The cycle strategy:** prisoner k opens box k first. It contains some number j ; he opens box j next; reads m there, opens box m — following the pointers. That is the entire strategy.

Why it works: box-to-content is a permutation, and every permutation decomposes into cycles. Starting at box k and following pointers walks exactly the cycle containing k , and that walk is guaranteed to reach the box holding the number k — the pointer that closes the loop. So prisoner k succeeds precisely when *his cycle is no longer than fifty*. His fate stops being fifty independent gambles and becomes one question about the permutation's shape.

With that, the group's fate collapses onto a single global event: **all hundred succeed if and only if the permutation contains no cycle longer than 50**. How likely is a long cycle? For any length $k > 50$, at most one such cycle can exist (two would need more than 100 elements), and the probability that a random permutation of 100 contains a k -cycle is exactly $1/k$. So the probability of failure is $1/51 + 1/52 + \dots + 1/100 = H_{100} - H_{50} \approx \ln 2 \approx 0.693$, and survival is $1 - (H_{100} - H_{50}) = \mathbf{0.31183}$, computed exactly. The number barely degrades as the prison grows: the limit is $1 - \ln 2 \approx 30.7$ percent, whether the prisoners number one hundred or one billion.

Now the part to sit with. Ask any single prisoner what the strategy did for *him*: the cycle through his box is equally likely to have any length from 1 to 100, so his personal success probability is $50/100$ — exactly $1/2$, identical to random guessing. No individual's odds improved at all. What changed is the *correlation*. Random guessers fail independently, so someone almost surely fails; cycle-followers fail together or not at all, because every member of a too-long cycle fails at once and, when no long cycle exists, nobody does. The strategy herds a hundred fates onto one coin flip about the shape of a

single permutation. And it is essentially unbeatable: Curtin and Warshauer (2006), building on the original analysis of Miltersen and Gál, showed no strategy does meaningfully better.

The problem was born in theoretical computer science — Peter Bro Miltersen posed it (with Anna Gál) at ICALP 2003 as a gadget in a data-structure lower bound, and his colleague Sven Skyum found the cycle strategy. It has been called, credibly, the most counterintuitive result in recreational mathematics.

The Move — Correlate your fates. When independent successes multiply into doom, redesign so that everyone succeeds or fails on the *same* event. A second move rode along: impose structure on randomness — the permutation's cycle decomposition was sitting there all along, waiting for someone to read boxes as pointers.

★★★★★ 10 · The Blue-Eyed Islanders

An island of two hundred perfect logicians: one hundred blue-eyed, one hundred brown-eyed. There are no mirrors and no reflective water, and eye color is never discussed — a taboo with teeth. Everyone sees everyone else's eyes; no one knows their own. The law: on the midnight after you *know* your own eyes are blue, you must leave the island forever. All of this — the law, everyone's perfect logic — is common knowledge. One day a visitor addresses the entire island at once: "I see at least one person with blue eyes." Every islander could already see at least ninety-nine blue-eyed people; the visitor seems to have said nothing new. What happens?

What did the visitor actually add? Sit with that.

Solution. All one hundred blue-eyed islanders leave together, at midnight of the hundredth day.

Climb from the bottom. One blue-eyed islander: she sees no blue eyes anywhere, so the announcement can only mean her. She leaves the first midnight. Two: each sees exactly one blue-eyed person and reasons, "if my eyes

are brown, he is alone — he will work it out and leave tonight." Nobody leaves, because each was waiting on the other. That silence is loud: each now knows there is a second pair of blue eyes, and whose. Both leave the second midnight. Three: each blue-eyed islander sees two and reasons, "if I am brown, those two resolve each other by night two." Midnight two passes, full island. All three know; they leave night three. The induction is on rails: n blue-eyed islanders leave on midnight n . With one hundred — ninety-nine silent midnights, then a mass departure on the hundredth.

But what did the visitor *add*? Everyone already saw blue eyes. Here is the honest answer, and it is the entire puzzle: before the announcement, "someone has blue eyes" was known to everyone — but it was not **common knowledge**. Follow the nesting. A blue-eyed islander sees 99 others. He knows each of those sees at least 98. He knows they know that the others see at least 97... Each level of "everyone knows that everyone knows" burns one islander of margin, and at depth 99 the chain bottoms out in a hypothetical islander who — inside all those nested imaginations — might see no blue eyes at all. The 99-deep statement was *false*. The visitor's announcement — heard by all, seen by all to be heard by all — makes "someone has blue eyes" true at every depth simultaneously. That is what was new: not a fact about eyes, a fact about knowledge of knowledge.

From then on, each silent midnight is itself a public announcement — "nobody yet knew" — and the induction climbs one rung per night until it reaches the truth. A public statement can change the world while adding zero first-order information. The puzzle descends from Littlewood's dirty-faces problem (*A Mathematician's Miscellany*, 1953) and Gamow and Stern's forty unfaithful wives (*Puzzle-Math*, 1958); the blue-eyed dress of it spread through the 2000s, notably via xkcd and Terence Tao's 2008 blog treatment.

The Move — Induct on common knowledge. "Everyone knows X" and "everyone knows that everyone knows X..." are different states of the world, and protocols, markets, and standoffs run on the difference. When a public event changes behavior without changing anyone's private facts, look for the depth of knowledge it completed.

★★★★★ 49 • Freudenthal's Impossible Puzzle

Two integers x and y are chosen, with $1 < x < y$ and $x + y \leq 100$. Mr. P is told only their product; Mr. S is told only their sum. The following conversation takes place, every word truthful and heard by both:

P: "I don't know the numbers." S: "I knew you didn't." P: "Now I know them."

S: "Now I know them too."

Find x and y — from the dialogue alone.

It has been called the impossible problem. It isn't — quite.

Solution. Not one word refers to the numbers, yet each line is a hard fact about what a speaker's information allows — and each fact filters the candidate pairs.

Line 1 — P doesn't know. His product must factor ambiguously within the constraints. Had it been, say, $35 = 5 \times 7$, two primes, he would have known at once. Every uniquely-factorable product dies here.

Line 2 — S *knew* P didn't. Not guessed: knew, from the sum alone. So **every** admissible split of S's sum must give an ambiguous product. This is savage. Any sum splittable into two primes fails — that split's product would have told P everything — and at this size every even sum can be so split (checked directly; no conjecture needed). Sieving the odd survivors leaves exactly ten sums: **{11, 17, 23, 27, 29, 35, 37, 41, 47, 53}**.

Line 3 — now P knows. Learning line 2 told P the sum lies in that ten-element list; his product must be compatible with *exactly one* member. Watch it work on the eventual answer: product $52 = 4 \times 13$, sum 17 — in the list; or 2×26 , sum 28 — even, out. One survivor. P announces knowledge.

Line 4 — now S knows too. Among all splits of his sum, exactly one can have become P-solvable at line 3. Run his sum, 17: the split (2,15) has product $30 = 2 \times 15$ (sum 17, listed) but also 5×6 (sum 11, also listed) — that P would still be stuck, so (2,15) is eliminated. (3,14): product 42, ambiguous between sums 17 and 23. Out. (5,12): 60, torn between 17 and 23. (7,10): 70, torn between 17 and 37. (8,9): 72, torn between 17 and 27. (6,11): 66, torn between 17 and 35. Only

(4,13) survives — so S, holding 17, now knows. And machine verification closes the loop: for every *other* sum in the list, two or more splits survive line 3, so line 4 would have been false. The sum is 17, the product 52: $x = 4$, $y = 13$.

A dialogue that looked content-free determined two numbers uniquely. Chapter 5's Cheryl's Birthday is this puzzle's pop-viral descendant, running the same engine at lower voltage. Hans Freudenthal published the original in *Nieuw Archief voor Wiskunde*, 1969; Gardner anglicized it a decade later as "the impossible problem" — impossible-seeming, because declared ignorance looks like the absence of information.

The Move — Mine the meta-statements. "I don't know" and "I knew you didn't" are mathematical facts about information sets — filters as sharp as equations. When a negotiator, a market, or a machine announces what it *cannot yet conclude*, that announcement is data. Nothing was said about the numbers, and that was the data.

★★★★ 50 • The Unexpected Hanging

A judge decrees on Sunday: "You will be hanged at noon on one weekday this coming week — and the hanging will be a surprise. You will not know the day beforehand." The prisoner reasons: "They cannot hang me Friday: by Thursday night, with four days gone, I would know it was coming. Friday is out. But with Friday struck, Thursday becomes the last possible day, so by Wednesday night I would know — Thursday is out too." Day by day the week collapses, and the prisoner concludes, relieved, that the sentence cannot be carried out at all. On Wednesday at noon the hangman arrives. The prisoner is completely surprised. The decree, in the end, was true. Where is the flaw in his reasoning?

Nearly eighty years of philosophy has not settled this one.

Solution. There isn't one — not a settled one — and this book will not pretend otherwise. What is agreed: the prisoner's elimination argument and the judge's eventual vindication cannot both be sound, and the fault must

sit somewhere in the reasoning about *knowing*. Where, exactly, remains contested across more than fifty academic papers.

Three families of diagnosis, none of them consensus. Quine (1953) attacked the base of the induction: eliminating Friday requires the decree to be not merely true but *knowably reliable at every future point* — and "the decree will be fulfilled" versus "I may lean on the decree inside my own reasoning" are different claims, silently conflated. The epistemic-logic school formalizes "surprise" with knowledge operators and finds the announcement self-referentially unstable: a statement that can be true yet not knowably true, so that the prisoner's own use of it destroys it — Sorensen (1988) calls the final-day case a "blindspot." The deflationary school shrugs: for a one-day week the decree is flatly inconsistent, for five days it is *approximately* assertable, and intuition smears the difference into an aura of paradox.

The problem reportedly began life as a Swedish wartime civil-defense announcement — a drill to be held next week, on a day no one could predict — and was chewed over until D. J. O'Connor put it into *Mind* in 1948. Gardner's 1963 column made it famous; Sorensen's *Blindspots* (1988) gave it a book. The ledger stays open.

Why end the boss room on an unbeaten boss? Because the genre should be honest about its edges, and because the lesson survives the missing solution. Every other puzzle in this room ran on knowledge behaving lawfully — counted through a bulb, nested on an island, mined from a dialogue. Here is the door marked *knowledge about one's own future knowledge*, and behind it the floor is missing. The honest response to an open problem is the sentence this book has tried to earn the right to say plainly: nobody knows yet.

The Move — Knowledge can self-destruct. Statements about what you *will know* are not ordinary facts; feeding them into your own inference can invalidate them. Respect the type-difference between what is true and what is knowably true — announcements of the form "you will be audited, unannounced" carry real weight in law and security, and this puzzle is the standing proof that they can be built on air.

The moves you now own

- **Maximize information per test** (★★★★ 8, the twelve balls) — budget outcomes first ($3^3 \geq 24$), then make every test split the survivors evenly.
- **Solve the endgame first** (★★★★ 44, the pirates) — anchor where the game is trivial; roll backward, pricing each vote at its alternative future.
- **Design a one-bit protocol** (★★★★ 9, the light bulb) — tiny shared state calls for a pre-agreed protocol, not deduction; appoint an aggregator.
- **Correlate your fates** (★★★★ 37, the boxes) — when independent success is hopeless, hang everyone's success on one shared event.
- **Induct on common knowledge** (★★★★ 10, the islanders) — "everyone knows" and "everyone knows everyone knows" are different worlds.
- **Mine the meta-statements** (★★★★ 49, Freudenthal) — declared ignorance is data; filter on information sets.
- **Knowledge can self-destruct** (★★★★ 50, the hanging) — truths about future knowledge may not survive being used.

These seven are Part II's summit ridge. Beyond their individual moves, they train a temperament: patience with long chains that must not leak, respect for answers that offend intuition and are true anyway, and — once, at the end — the composure to leave a question open because it is open. Part III visits the people who climb grades like this for sport, and takes their training apart.

PART III

The Olympiad Mind

CHAPTER 8

How Champions Think

Competition mathematics looks like a test of raw cleverness. It is mostly a test of trained recognition. What separates an IMO medalist from a strong classroom student is not processor speed; it is a library of perhaps two hundred recognition-action pairs — "this problem smells like parity," "these constraints scream binary search" — compiled from thousands of solved and, crucially, failed problems. Nobody is born knowing that "minimize the maximum" is a sentence about binary search. Someone drilled it, missed it, journaled the miss, and met it again.

The psychology behind this is old and solid. Adriaan de Groot's chess studies of the 1940s, extended by Chase and Simon's "Perception in Chess" (1973), showed that grandmasters do not search deeper than club players in most positions. They see better. In the classic experiment, a real position flashed for a few seconds is rebuilt almost perfectly by a master and haltingly by a novice — but scramble the pieces into a random arrangement and the master's advantage nearly vanishes. Years of study have chunked the board into meaningful patterns, and expertise converts search into lookup. Chase and Simon put the master's vocabulary at tens of thousands of chunks — the size of a reading vocabulary, which is the right comparison: strong players read positions rather than compute them. A heuristic is compressed experience: it prunes an exponential tree of possibilities down to a handful of candidate moves.

This chapter is about the mind that carries such a library: the frame it thinks in, the flaw in the frame, the culture that builds the library, and the craft of reading a problem's author like a dossier. It ends with the machines that started winning the same medals in 2024–25, and with the reason the training culture did not thereby become a museum. The next three chapters are the library itself — sixteen tools, each with its recognition signal and a worked pair. This chapter is the operating manual for the librarian.

The four phases

George Pólya, a Hungarian mathematician at Stanford, published *How to Solve It* in 1945. It has sold over a million copies and is the founding document of heuristics-as-curriculum — the first sustained argument that problem-solving is teachable technique rather than temperament. Everything in Part III hangs on his frame, so it is worth seeing whole before the tools arrive.

Phase 1 — Understand the problem. Pólya's questions: What is the unknown? What are the data? What is the condition? Is the condition even sufficient to determine the unknown? Draw a figure. Introduce suitable notation. The competition translation: restate the problem in your own formalism before touching it. Most wrong solutions die here — they solve a neighboring problem, often perfectly. ICPC teams ritualize this phase: one member re-reads the statement hunting for the sentence everyone skimmed.

Phase 2 — Devise a plan. Have you seen it before? Do you know a related problem? Can you restate it? Can you solve a simpler or analogous problem first? Have you used all the data? This phase is the arsenal of Chapters 9 through 11. "Devising a plan" means running your recognition signals over the problem until one fires.

Phase 3 — Carry out the plan. Check each step; can you see clearly that the step is correct — can you prove it? Translation: execution discipline. Case-work bookkeeping, algebra hygiene, off-by-one vigilance, and, in programming contests, the test-before-submit ritual. Plans are cheap. Flawless execution under a clock is a separately trained muscle, and contests are lost in this phase at least as often as in the previous one. ICPC teams debug printed code on paper while a teammate types the next solution — execution discipline made spatial.

Phase 4 — Look back. Can you check the result? Derive it differently? Use the method somewhere else? This is the phase amateurs skip and champions industrialize. The upsolving culture described below is Phase 4 turned into a lifestyle.

Two notes before the caveat. The phases loop rather than march: a failed plan sends you back to understanding, and looking back at a solved sub-problem often is the plan for the main one. And Phase 2 is where the next three chapters live — Chapter 9's kill shots, Chapter 10's builders, and Chapter 11's lenses are the plan-devising phase unpacked into sixteen named reflexes.

The missing dial

The caveat that keeps the frame honest comes from Alan Schoenfeld's *Mathematical Problem Solving* (1985). Schoenfeld watched students who had been taught Pólya's heuristics fail anyway — not for lack of tools but for lack of *control*. His timeline charts of student sessions are famous in the field: read the problem, pick the first idea that offers itself, chase it in a flat line until the bell. An expert's chart oscillates — analyze, plan, try, check, back out, replan. The students owned tools; they lacked the timer in the head that periodically interrupts with *is this approach earning its keep?* Schoenfeld's classroom fix was almost comically simple: he interrupted working students, again and again, with three questions — *What exactly are you doing? Why are you doing it? How does it help you?* — until the interruption installed itself and students began asking unprompted.

Call metacognitive monitoring the invisible seventeenth heuristic. Tools choose moves; control decides when to abandon them. Two minutes of algebra with no simplification — reconsider. A case split that has bred seven sub-cases — back out and hunt for an invariant instead. Control is also the skill the clock trains hardest: under contest pressure, sunk cost whispers *five more minutes* precisely when the dial says switch. Everything in the next three chapters assumes that loop is running, because a library without a librarian is just a pile.

Where the library comes from

The public imagines prodigies. The community runs a training industry, with unusually honest feedback loops.

The arenas. Each major arena drills a different muscle, and serious competitors cross-train. The International Mathematical Olympiad, held since 1959 and now drawing over a hundred countries: two days, three problems per day, four and a half hours a day, seven points per problem, 42 the perfect score. Roughly, problems 1 and 4 are entry, 2 and 5 middle, 3 and 6 brutal. The International Olympiad in Informatics is its high-school programming twin — partial scoring, two five-hour days. ICPC, the university team championship, seats three students at one shared computer for five hours and ten-to-thirteen problems; the engineered scarcity forces role specialization — reader, coder, prover — and paper-debugging, which trains something subtle: the verbalization of half-formed thought. The Putnam, running since 1938, offers twelve problems for 120 points and a median score that in many years is 0 or 1 — a calibration that normalizes failure as the default state of problem solving. The calibration is deliberate and pedagogically load-bearing: a sport where the median entrant scores 1 out of 120 cannot be about feeling clever, so it becomes about the long game of getting less wrong. Codeforces, founded in 2010 by Mike Mirzayanov in Saratov, is the sport's daily gym: rated two-hour rounds and a ladder of colors from gray newbie through cyan, blue, purple, and orange to red grandmaster, with a black-and-red "legendary" tier past 3000. Gennady Korotkevich — "tourist," the sport's Federer — became the first to cross rating 4000 in 2024. On the mathematical side, Art of Problem Solving, founded in 2003 by Richard Rusczyk, is forum, textbook press, and nursery; most recent US IMO team members grew up on it.

Upsolving, the core loop. The single most agreed-upon practice in the culture: after every contest, solve the problems you failed — that day, before the sting fades. Rating measures the past; upsolving builds the future. A missed problem marks the precise boundary of your ability, and converting it from failure into owned pattern moves the boundary — the contest already found your weakest spot for free, and upsolving is refusing to waste the diagnosis. Competitors speak of a problem "entering the hands" only once it has been re-solved from scratch. Understood-from-the-editorial does not count: reading a solution feels like understanding, but only reproduction from a blank page proves the pattern is installed rather than admired.

The mistake journal. The stronger version, kept by many high-rated competitors: every wrong submission classified — misread statement, wrong algorithm, off-by-one, unproven greedy, overflow — and every failed problem logged, then revisited on a schedule and re-solved cold. This is spaced repetition applied to reasoning patterns instead of vocabulary; the deck is your personal error distribution. The classification step matters as much as the review. "I lose most of my rating to unproven greedys" is an actionable diagnosis, and Chapter 10 prescribes for exactly that disease. The journal's power is the denominator: after fifty entries you know your base rates, and base rates turn vague resolve into a targeted drill.

Ladders and the +200 rule. Practice slightly above your level — the folk prescription is about 200 rating points up: hard enough to hurt, close enough to finish. Ahmed Aly's A2OJ "ladders" formalized the idea as ordered problem lists; per-problem difficulty ratings and curricula like the USACO Guide carry it today. The logic is deliberate practice in Ericsson's textbook sense — maximal time at the edge of competence, minimal time in comfort. Below the edge nothing is learned; too far above it nothing is finished, and finishing is where the pattern sets.

Struggle before editorial. A community norm with real disagreement about the number and none about the principle: fight a problem genuinely — thirty to sixty minutes of motion, not stall — before reading its editorial; then read one hint's worth and try again. Reading solutions too early trains recognition of solutions, not production of them. The struggle is not the price of the learning. It is the learning: the recognition signals of the coming chapters get wired during the fight.

Stress testing. The competitive programmer's version of the scientific method: when a solution is suspect, write a five-line brute force, a random test generator, and a loop that diffs the two until a minimal counterexample falls out. Strong competitors do this *during* contests; it converts "I believe my greedy" into evidence within a minute. Note the shape of the habit — mechanical, adversarial, aimed at yourself. It transfers directly to auditing the output of machines, a point Part V will insist on.

Virtual contests. Re-run old contests under a real clock, full rules. This trains the meta-skills no problem list teaches: reading all the problems first, ordering them by expected value, deciding when to abandon an approach — Schoenfeld's dial, drilled deliberately — and managing the panic of minute 100. ICPC teams rehearse the logistics of who touches the keyboard when, like a pit crew. None of it is glamorous, and that is the point: the community treats reasoning as a trainable motor skill — with a gym, a technique, and a log book.

The canon. Pólya (1945). Engel's *Problem-Solving Strategies* (1998), with hundreds of drills per tool. Zeitz's *The Art and Craft of Problem Solving*. The AoPS book series, and the IMO Shortlists, published a year after each olympiad — the community's graded problem bank and every national team's training set. For programming: Laaksonen's *Competitive Programmer's Handbook*, Halim's *Competitive Programming*, and cp-algorithms.com as the working reference.

The setter's mind

Problems are not found. They are manufactured — and knowing the factory's methods is itself a solving tool, the way a locksmith picks locks faster for knowing how locks are cut.

How the sausage is made. Three patterns dominate. *Kernel-first*: the setter starts from a phenomenon — an invariant that surprises, a greedy that works for a non-obvious reason, a bijection between unlike objects — then wraps a story around it and tunes constraints until the intended idea is both necessary and sufficient. The story, whether knights or factories or chocolate bars, is deliberately load-free costume. *Backwards-first*: the setter fixes the shape of the answer — "the extremal configuration should be the golden-ratio pair" — and constructs a rule set for which that is the solution. Problem design as inverse problem solving. *Kill the unintended*: the drafting loop is adversarial. Testers attack with naive greedys, brute force, and pattern-guessing, and every successful cheap attack gets patched — a constraint raised to kill the brute force, a test added to break the plausible-but-wrong greedy, sample cases curated not to reveal the trick. If a plausible greedy survives testing, the setter builds the case where it fails —

remember that when the $\{1, 3, 4\}$ coins ambush you in Chapter 10. Codeforces problems pass through Polygon, a preparation system of validators, checkers, generators, and stress tests: industrial QA for puzzles. At the IMO, countries submit proposals, a Problem Selection Committee distills a short-list of around thirty across algebra, combinatorics, geometry, and number theory, and an international Jury picks and polishes the final six.

The three factory patterns matter because each leaves fingerprints on the finished problem — and fingerprints can be read. Therefore read every problem as a message from its maker.

- **Constraints leak the intended complexity.** In a programming contest, $n \leq 20$ whispers bitmask; $n \leq 40$ shouts meet-in-the-middle; $n \leq 200,000$ with sums says $n \log n$. The full decoder table waits in Chapter 11 — read in reverse, it is a decoder of authorial intent.
- **Every hypothesis is used.** Olympiad statements are minimal by editing pressure. If it says *distinct* integers, distinctness is load-bearing; if the graph is *connected*, the proof will need it. A condition you have not used is either the key you are missing or the door you have not found — and, contrapositively, a "solution" that never used the strange condition is almost surely wrong.
- **Strange exactness is a signpost.** "Exactly $2n - 1$ numbers"; " $n^2 + 1$ terms." Sharp-looking counts mean the bound is tight, and the example showing tightness is part of the intended understanding. Find it and you have usually found the mechanism. Both of those particular counts will detonate in the coming chapters.
- **Position is prior.** Problem 1 of an IMO day has a short, honest solution — if yours is a monster, restart. A mid-division contest problem wants one idea plus clean execution, not a research program. The setter's calibration is free Bayesian evidence, and strong solvers consciously condition on it. The conditioning works in reverse too: a trivial-looking Problem 6 is a warning, not a gift.
- **Cultural fingerprints.** A modulus of 998244353 — a prime p with 2^{23} dividing $p - 1$ — announces number-theoretic-transform convolution to the initiated, the way $10^9 + 7$ announces "just avoid overflow." These tells

are not trivia; they are the strongest single hint on the page, because they encode what the author expects the solution to *be*. Sample outputs that hide the interesting case are telling you the interesting case exists.

The deeper point, for a book about thinking: expert solvers run a theory of mind on the adversary-author. *Why would a sane designer include this clause?* is not mathematics at all — it is hermeneutics. And it is precisely the skill that transfers out of the arena, to contracts, regulations, benchmarks, and AI evaluations. Artifacts built by intent are decoded fastest by modeling the intent.

The machines arrive

An honest chapter about olympiad minds has to end with the news. In July 2024, DeepMind's AlphaProof and AlphaGeometry 2 reached silver-medal level on that year's IMO problems — 28 of 42 points, one point below the gold cutoff — with an asterisk: the geometry problem fell in seconds, but some of the rest took days, well beyond contest conditions. A year later the asterisk was gone. In July 2025, Gemini Deep Think was officially graded gold-medal standard at IMO 2025 — 35 of 42, inside the four-and-a-half-hour limits, working in natural language — and OpenAI announced a comparable experimental result the same week. As of mid-2026, frontier models also perform at grandmaster-plus level on competitive-programming benchmarks. Every claim in this paragraph is dated on purpose; in this genre, undated claims are already wrong.

So a fair question now hangs over everything this chapter described: why train a human to do what a data center medals in?

Hold the question; it belongs to Part IV, where it gets dates, numbers, and the occasional surprise rather than a shrug. But two things can be said already, because they follow from this chapter alone. First, the culture was never really about the medals. It is the most refined curriculum humanity has built for installing recognition primitives in a human head, and the primitives, not the trophies, are what transfer. Second, look back at which habits the culture drills hardest: stress-testing your own solution, journaling your own errors, struggling before reading answers, decoding a hidden

author's intent. A mind trained this way is not merely a generator of solutions. It is an auditor of them — anyone's, including a machine's.

The arsenal itself comes next: five kill shots, five building tools, six ways of seeing — each with its definition, its recognition signal, a worked pair of problems, and an honest line on what machines can and cannot yet do with it. What you have just read is how to hold them.

CHAPTER 9

The Kill Shots

Some tools build a solution piece by piece. These five end the fight outright. A kill shot is a single observation that makes an exponential search unnecessary: one line where a machine, working without the abstraction, would drown in cases.

Each tool gets the same treatment: a definition, which every textbook has; a recognition signal, which textbooks leave tacit — the *smell* a trained mind notices, often pre-verbally, before the hand reaches for the tool; a worked pair, one teachable in a sitting (★★), one at genuine olympiad level (★★★★); and a line on how machines fare. The signals are the real payload. Learn the smell, and the tool fires itself.

Pigeonhole

The tool. Put more than n objects into n boxes and some box holds at least two. In general, n objects in k boxes put at least $\lceil n/k \rceil$ into some box. That is the entire theorem. Its power lives in the choice of boxes — which is never given in the problem. It must be invented.

The smell. "Prove there exist two..." or "show that some..." with no obvious construction. A quantity bound: "among any 51 numbers...", "any five points in a square...". A guarantee demanded no matter how an adversary arranges things. The trained reflex: *what finite set of categories can I force these objects into?* Distances suggest geometric cells; divisibility suggests residues or odd parts; sequences suggest prefix sums. Two Londoners have the same number of head hairs — nine million people, at most about a million possible hair counts; the boxes are the counts. Any 100 integers contain a consecutive block summing to a multiple of 100 — the boxes are the residues of the 101 prefix sums $S_0, \dots, S_{100}$; two collide, and the block between them is their difference. Aimed at the continuum, the same one-liner invents ra-

tional approximation: Dirichlet's theorem — for any real α some $q \leq N$ has $|q\alpha - p| < 1/N$ — is $N + 1$ fractional parts falling into N intervals.

★★ • Five Points in a Square

Prove that among any five points in a closed unit square, two lie at distance at most $\sqrt{2}/2 \approx 0.707$.

Pause here. The page can wait.

Solution. Cut the square into four $\frac{1}{2} \times \frac{1}{2}$ quadrants. Five points, four quadrants: some quadrant holds two of them. Two points of a $\frac{1}{2} \times \frac{1}{2}$ square are at most a diagonal apart: $\sqrt{((\frac{1}{2})^2 + (\frac{1}{2})^2)} = \sqrt{2}/2$. ■

Nothing in the statement mentions quadrants. The solver manufactures the boxes so that "same box" forces the conclusion — and the choice of four boxes, not nine and not two, is reverse-engineered from the number five: one box fewer than the points.

The Move — Manufacture the boxes. The theorem is free; the boxes are the work. Read the conclusion you want, then invent categories in which a collision delivers it.

★★★★ • The Unavoidable Run

Prove that every sequence of $n^2 + 1$ distinct real numbers contains a monotone subsequence — increasing or decreasing — of length $n + 1$.

Try it before reading on. The boxes are well hidden.

Solution. To each position i attach a two-number label (x_i, y_i) : the length of the longest increasing subsequence ending at i , and of the longest decreasing one ending there. Suppose no monotone subsequence reaches length $n + 1$. Then every x_i and y_i lies in $\{1, \dots, n\}$: at most n^2 distinct labels for $n^2 + 1$ positions. Pigeonhole: two positions $i < j$ share a label. But the numbers are distinct, so either $a_i < a_j$ — and any longest increasing run ending at i ex-

tends by a_j , forcing $x_j \geq x_i + 1$ — or $a_i > a_j$, forcing $y_j \geq y_i + 1$. Either way the labels differ. Contradiction. ■

This is Erdős–Szekeres (1935), carrying the setter's tell from Chapter 8: " $n^2 + 1$ terms" is strange exactness — n^2 terms can dodge, via n descending blocks of n , each block higher than the last. The boxes here are pairs of statistics *about the object itself*: nobody "sees" n^2 boxes in the statement; the solver builds a two-coordinate fingerprint precisely so that a collision is impossible, then counts fingerprints. A companion classic on the same fuel: any $n + 1$ numbers from $\{1, \dots, 2n\}$ include one dividing another — write each as $2^k \cdot m$ with m odd; only n odd parts exist; two numbers share one, and the smaller-powered divides the other.

The Move — Fingerprint and collide. Attach monotone statistics until a collision would be absurd; then force the collision by counting. The double-label trick recurs all over combinatorics.

Machines and this move. Verifying a pigeonhole proof is trivial, but finding the boxes is invention over an infinite space of partitions — and Haken proved in 1985 that resolution proof systems, the engine inside classical SAT solvers, need exponentially long proofs of the pigeonhole formula itself. One human abstraction, one exponential wall bypassed.

Invariants & Monovariants

The tool. An invariant is a quantity no legal move can change — a conserved charge. To prove a target unreachable, exhibit an invariant that differs between start and target. A monovariant moves in one direction only; use it to prove a process must stop, and to see where it can end. This is the move that killed the 14–15 swap in Chapter 2: permutation parity times blank-row parity never changes, so Loyd's thousand-dollar prize was never in danger.

The smell. The problem describes moves — "erase two numbers and write...", "two chameleons meet and...", "flip any row" — and asks *can you reach this? does it stop? what are the possible ends?* The reflex chain: compute a few cheap quantities (sum, sum mod 2, mod 3, sign of the product, inversion count) before and after one generic move; whatever survives un-

changed is your candidate. For termination, hunt a bounded quantity forced monotone. The deep habit: ignore the trajectory entirely and study what the dynamics conserve — Noether's instinct in miniature.

★★ • The Blackboard Reduction

The numbers 1 through 10 stand on a blackboard. A move: erase any two of them, a and b , and write $|a - b|$. After nine moves a single number remains. Can it be 0?

Pause. Play a few moves in your head first.

Solution. Track S , the sum of everything on the board, mod 2. A move changes the sum by $(a + b) - |a - b| = 2 \cdot \min(a, b)$ — an even number — so $S \bmod 2$ never changes. It starts at $1 + 2 + \dots + 10 = 55$, odd. The final number is therefore odd, and in particular never 0. ■ (Odd endings are achievable down to 1; the invariant says exactly which endings are barred.)

The reachability graph of this game has thousands of states. The invariant collapses it to two.

The Move — Find the conserved charge. Decide reachability without exploring a single branch: compute what every move preserves, then compare start with target.

★★★★ • The Pentagon That Must Stop

(IMO 1986, Problem 3.) Integers $x_1, \dots, x_5$ sit at the corners of a pentagon, with total sum $s > 0$. Whenever some corner value y is negative, you may replace its consecutive triple (x, y, z) by $(x + y, -y, y + z)$. Prove that this procedure necessarily terminates.

Pause here — check first what a move leaves alone.

Solution. The move conserves the total: $(x + y) + (-y) + (y + z) = x + y + z$. So the sum stays $s > 0$ forever. Now define $f(x_1, \dots, x_5) = \sum_i (x_i - x_{i+2})^2$, indices mod 5 — a nonnegative integer. Apply the move at position 4 (cyclic sym-

metry makes that typical), with $y = x_4 < 0$. Four of f 's five terms change; $(x_3 - x_5)^2$ survives, since both its endpoints shift by y . Expanding the four affected squares and cancelling:

$$\Delta f = [-2y(x_1 - x_3) + y^2] + 4x_2y + 4x_1y + [2y(x_5 - x_2) + y^2] = 2y(x_1 + x_2 + x_3 + x_5) + 2y^2 = 2y \cdot s.$$

Since $y < 0 < s$, every move strictly decreases f . A strictly decreasing nonnegative integer allows only finitely many moves. ■

Now the honest part: nobody derives f by pure taste in an exam room; you hunt. You want a nonnegative integer that provably drops; sums and linear forms are conserved — useless — so you try the natural quadratic forms, Σx_i^2 , $\Sigma(x_i - x_{i+1})^2$, $\Sigma(x_i - x_{i+2})^2$, computing Δ for each until one factors through s . In 1986, only eleven students solved this problem. The gap between *checking* an invariant and *inventing* one is the gap between ★★ and ★★★★★.

A companion drill: 13 red, 15 green, 17 blue chameleons; when two of different colors meet, both turn the third color. All-one-color is impossible — every meeting shifts each pairwise difference of counts by 0 or ± 3 , so the differences mod 3 are conserved; they start at 2, 2, 1, and a monochrome state needs a 0.

The Move — When nothing is conserved, find what only falls. Termination wants a bounded monotone quantity. Hunt it in a parametrized family and let the algebra pick the member.

Machines and this move. Checking an invariant is one line of algebra; discovering one is synthesis over an infinite function space — the same problem, called loop-invariant inference, is still a benchmark bottleneck in program verification, and language models do well on famous invariants and poorly on freshly minted dynamics.

The Extremal Principle

The tool. Interrogate an extreme witness: the largest, smallest, first, closest, boundary case. Mode one — structure: the extremal object's extremality

forbids certain configurations around it, which is exactly the leverage you need. Mode two — minimal counterexample, Fermat's *descente infinie*: assume a counterexample exists, take the smallest, construct a smaller one; therefore none exists. Descent in one breath: if $\sqrt{2} = a/b$ with b minimal, then $a^2 = 2b^2$ forces $a = 2c$, whence $2c^2 = b^2$ — and (b, c) is a strictly smaller representation. Dead. The same skeleton disposes of $x^4 + y^4 = z^4$, Fermat's own showcase.

The smell. "There exists an element with property P " in a finite mess — first ask whether the maximum or minimum has P for free. Any configuration where "pick the worst offender and improve it" seems available. Geometry with closest pairs, smallest circles, minimal angles. Before inventing machinery, ask what the extremal element already knows: an extremum ships with a free contradiction generator, since anything more extreme than it is impossible by fiat.

★★ • The Tournament King

In a round-robin tournament — every pair plays once, no draws — prove there is a player v such that every other player either lost to v , or lost to someone who lost to v .

Pause. Whom would you nominate?

Solution. Nominate v , a player with the maximum number of wins. Suppose some u is not reachable in two steps: then u beat v , and u also beat every player v beat — for if v beat w and w beat u , u would be reachable. So u 's victims include all of v 's victims plus v himself: $\text{wins}(u) \geq \text{wins}(v) + 1$, contradicting maximality. ■

No induction, no construction, no algorithm. The maximal element is *forced* to have the property, because a violation would manufacture something greater than the maximum.

The Move — Ask what the maximum already knows. One extremal choice replaces an entire search.

★★★★ · Sylvester's Line

A finite set of points in the plane is not all on one line. Prove that some line through two of the points contains no third point of the set. (Sylvester posed it in 1893; it stood open for about forty years. The proof below is L. M. Kelly's — widely called a proof from The Book.)

Pause. What would you minimize?

Solution. Consider all pairs (P, ℓ) where ℓ is a line through at least two points of the set and P is a point of the set not on ℓ . Such pairs exist, since the set is not collinear, and there are finitely many — so choose the pair minimizing the distance from P to ℓ . Claim: ℓ carries exactly two points. Suppose it carried three. Drop the perpendicular from P to ℓ , with foot F . Of the three points, two lie on the same closed side of F : call them A (nearer F , possibly at F) and B (farther), so the order along ℓ is F, A, B with $A \neq B$. A is not on the line PB — otherwise PB would meet ℓ twice and P would lie on ℓ . Now double-count the area of triangle ABP : $\frac{1}{2} \cdot |AB| \cdot d(P, \ell) = \frac{1}{2} \cdot |PB| \cdot d(A, PB)$, so $d(A, PB) = d(P, \ell) \cdot |AB| / |PB|$. But $|PB| = \sqrt{PF^2 + FB^2} > FB \geq AB$, since P is off the line ($PF > 0$) and B lies beyond A ($FB \geq AB$). Hence $d(A, PB) < d(P, \ell)$: a point-line pair closer than the minimum. Contradiction. ■

Forty years of failed attacks, then half a page whose only idea is *choosing what to minimize* — and the craft is the choice: a joint point-line object, neither alone. "How would I ever find this?" The honest answer is a checklist, run the way a chess player enumerates checks: closest pair? smallest angle? minimal area? minimal point-line distance?

The Move — Choose what to minimize. The extremal principle is only as strong as the functional you hand it. Enumerate candidates deliberately.

Machines and this move. Computing an argmax is instant; wielding extremality as a proof lever means inventing the thing to extremize — the pigeonhole-box bottleneck again; descent's orderings are the "termination measures" automated verifiers still need humans to supply.

Symmetry & Pairing

The tool. Exploit a transformation that preserves the problem. Three workhorse forms. *Pairing*: match objects so each pair carries a constant or cancelling contribution. *Mirroring*: in games, answer every opponent move through a symmetry, so the future is pre-answered. *Strategy stealing*: a symmetry between the players shows the first player's extra move can never hurt — so a first-player win exists, exhibited or not.

The smell. Terms in complementary pairs — k with $n + 1 - k$, a subset with its complement, a divisor d with n/d . A center, an axis, an involution: an operation that undoes itself. "Who wins with perfect play?" on a symmetric board. The reflex: hunt for an involution and ask what its fixed points are. Sums collapse; existence proofs fall out of the unpaired elements. (The most famous pairing in mathematics is a schoolroom legend about young Gauss; the story — and what is wrong with it — closes Chapter 11.)

★★ • Coins on a Round Table

Two players alternately place identical circular coins on a round table; coins may not overlap. Whoever cannot move loses. Who wins?

Pause. The table is round for a reason.

Solution. The first player — by seizing the one self-symmetric point. Move 1: a coin dead center. Thereafter, mirror: whenever the opponent plays at a spot S , reply at S' , its 180° rotation about the center. The invariant: after each of your moves, the whole configuration is symmetric under that rotation. So when the opponent finds any legal spot S , its mirror S' is empty — the position was symmetric before their move — and legal, since it clears the mirror image of everything S cleared; and $S' \neq S$, because the rotation's only fixed point, the center, was taken on move one. The table is finite, so someone runs out of moves. It is never you. ■

The mirroring player never calculates. The symmetry answers every future question in advance — a complete strategy compressed into one sentence.

The Move — Seize the fixed point, then mirror. Whoever owns the symmetry owns the endgame.

★★★ • Nim, or Mirrors Made of Bits

Three piles of stones: 3, 4, 5. A move takes any positive number of stones from any one pile. Taking the last stone wins. Who wins, and how?

Stop. Play the two-pile version first — say 2 and 2.

Solution. Two equal piles fall to pure mirroring: the second player repeats each move on the other pile, restoring (n, n) after every turn — a reply always exists, so the last stone is theirs. Unequal piles: the first player equalizes, then becomes the mirror.

Bouton's theorem (Charles Bouton, 1901 — the analysis that founded combinatorial game theory) generalizes the mirror to any number of piles. Write the pile sizes in binary. A position is lost for the player to move exactly when every binary column contains an even number of 1s — equivalently, when the XOR of the pile sizes, the *nim-sum*, is 0. Both directions take one observation each. From a zero position, any move edits a single pile, flipping at least one bit — some column goes odd; the balance always breaks. From a nonzero position, look at the highest unbalanced column; some pile has a 1 there; shrinking that pile to (itself XOR the *nim-sum*) rebalances every column — and it is a genuine shrink, because its leading change turns a 1 into a 0. The empty position has XOR 0, and its mover has already lost. So the winning policy is simply: always hand your opponent a zero.

Here: 3, 4, 5 = 011, 100, 101. The XOR is 010 $\neq$ 0, so the first player wins. The twos column is the unbalanced one, and only the 3-pile has a 1 there: play $3 \rightarrow 3 \oplus 2 = 1$. Check: 001, 100, 101 — every column even. It is the *only* winning opening; balancing through the 4-pile or 5-pile would require them to grow.



The equal piles were the special case the mirror explained: (n, n) is balanced column by column. Bouton's zero is mirroring generalized — when the

board offers no geometric symmetry, build an arithmetic one and keep it balanced.

The Move — Pair the bits. The involution can live in the arithmetic rather than on the board. Find the quantity whose balanced states you can always restore and your opponent can never preserve.

★★★★ • Chomp, and the Stolen Strategy

Chomp (David Gale, 1974): an $m \times n$ chocolate bar, $mn \geq 2$, with the bottom-left square poisoned. A move: choose any remaining square and eat it together with every square above and to the right of it. Whoever eats the poison loses. Prove the first player wins — for every board size — even though no general winning strategy is known.

Pause here. You will not find the strategy. That is the point.

Solution. Chomp is finite, deterministic, perfect-information, and cannot draw — so by backward induction (Zermelo), exactly one side owns a winning strategy. Suppose it is the second player. Let the first player open with the smallest possible nibble: the single top-right square. By assumption the second player has a winning reply Y , leading to some position Q whose mover — player one — loses. But observe: *every* Chomp move removes the top-right square anyway. So Q is exactly the position reached if Y had been played as an opening move. Player one could therefore have opened with Y , seating player two on the losing side of Q . Contradiction — so the winner is player one. ■ The proof names no strategy, and for general boards the winning first move is still unknown, computed only for small cases.

The symmetry here is between the *players*: anything your winning reply achieves, I could have achieved one move earlier. The same skeleton gives Nash's argument for Hex — an extra stone never hurts — and tic-tac-toe's "the first player does not lose."

The Move — Steal the strategy. Prove the win exists by pre-empting the second player's hypothetical strategy. Existence without exhibition.

Machines and this move. Engines detect and exploit board symmetries for pruning, but strategy stealing is a proof that needs zero search and yields zero policy — a reinforcement learner can master *playing* Chomp and still cannot say why player one wins.

Coloring & Parity

Chapter 6 ended on a broken chessboard: two opposite corners gone, thirty-one dominoes offered, and a two-sentence refusal — every domino covers one dark square and one light one, and the two missing corners shared a color. Take the generalization now, because it is a whole arsenal.

The tool: assign each cell a color or weight so that (a) every piece, wherever it lands, covers a fixed color pattern, while (b) the board's total color count is one the pieces cannot compose. A coloring is a homomorphism: it maps a vast space of attempted tilings onto tiny arithmetic — counts mod k — where the impossibility becomes visible. Parity arguments are the two-color special case; the whole device is this chapter's conserved charge specialized to tilings — an invariant wearing pigment.

The smell. Tiling and covering questions; traversals on grids — knight tours, lamp-flipping games; and above all this signature: the area arithmetic says *possible* while every attempted construction stubbornly fails. That stubbornness is the smell of a hidden obstruction. The reflex: try the chessboard coloring, then stripes, then diagonals mod k — with k matched to the tile's length. Tune the pigment's period to the piece's footprint.

★★ • Twenty-Five T's

Can twenty-five T-tetrominoes tile a 10×10 board? (Area check: $25 \times 4 = 100$. Arithmetic has no objection.)

Pause. Color first, then count.

Solution. Chessboard-color the board: 50 dark squares, 50 light. Wherever a T-tetromino lands, it covers three squares of one color and one of the other — an odd number of dark squares either way. Twenty-five T's therefore cover a total number of dark squares that is a sum of twenty-five odd numbers: odd. But the board has exactly 50 dark squares — even. Impossible. ■

The Move — Let parity see what area cannot. When the counts pass and the constructions keep failing, a two-coloring often knows why.

★★★★ · The Ten-by-Ten That Refuses

Prove that a 10×10 board cannot be tiled by twenty-five straight tetrominoes — 1×4 tiles. (The area check passes again. The trap is armed.)

Pause here. The chessboard pigment is too coarse for this one.

Solution. Color cell (i, j) — rows and columns counted from 0 — with $(i + j) \bmod 4$: diagonal stripes of period four. Key property: a straight tetromino, horizontal or vertical, occupies four cells whose $(i + j)$ values are consecutive — hence exactly one cell of each color 0, 1, 2, 3. So twenty-five tiles would cover each color exactly 25 times: a tiling forces 25 cells of every color. Now census the board. Among $i \in \{0, \dots, 9\}$, the residues 0 and 1 occur three times each ($\{0, 4, 8\}$ and $\{1, 5, 9\}$), residues 2 and 3 twice each; likewise for j . Counting cells with $(i + j) \equiv r$ means convolving $c = (3, 3, 2, 2)$ with itself:

- color 0: $3 \cdot 3 + 3 \cdot 2 + 2 \cdot 2 + 2 \cdot 3 = 25$
- color 1: $3 \cdot 3 + 3 \cdot 3 + 2 \cdot 2 + 2 \cdot 2 = 26$
- color 2: $3 \cdot 2 + 3 \cdot 3 + 2 \cdot 3 + 2 \cdot 2 = 25$
- color 3: $3 \cdot 2 + 3 \cdot 2 + 2 \cdot 3 + 2 \cdot 3 = 24$

(Sanity: $25 + 26 + 25 + 24 = 100$ ✓.) Colors 1 and 3 hold 26 and 24 cells — not 25. No tiling exists. ■

The coloring is engineered to the tile: period-four diagonals *because* the tile has length four and must eat one cell of each color. (The general theorem lurking behind the trick — de Bruijn, 1969: an $a \times b$ rectangle can be tiled by $1 \times n$ exactly when n divides a or b .) The ★★★★★ skill is not applying a coloring. It is designing one whose arithmetic the tile respects and the board violates.

One warning about scope: colorings prove only impossibility. Possibility needs a construction — and Chapter 10 opens with the construction machine.

The Move — Engineer the coloring to the tile. Match the pigment's period to the piece's footprint, then count what the board actually has.

Machines and this move. A non-tileable board is a SAT solver's worst case — John McCarthy circulated the mutilated chessboard as a "tough nut" for proof procedures — yet the coloring certificate, once conjectured, is machine-checkable in microseconds. Conjecturing it is the human part.

The five, at a glance

Move	The smell	Emblem
Manufacture the boxes	"prove two must share..."	five points, $\sqrt{2}/2$
Conserved charge / monovariant	moves given; "can you reach it? does it stop?"	the pentagon's $\Delta f = 2ys$
Interrogate the extremum	"some element must exist..."	the tournament king
Mirror, pair, steal	symmetric board; perfect play; paired terms	Nim's balanced columns
Color to contradict	area says yes, attempts say no	the ten-by-ten that refuses

Five ways to end a fight before it starts — each replacing a search with an observation, which is exactly why machines find them hardest to originate

and easiest to check. The next chapter's tools do the opposite kind of work. They build.

CHAPTER 10

The Builder's Tools

The kill shots end problems with one observation. These five tools do the opposite work: they construct — an answer, a tiling, a schedule, a strategy — piece by piece, with a proof riding along. Same format: definition, signal, worked pair, machines.

Induction, and the Art of Asking for More

The tool. Prove the base case; prove that each case forces the next. The non-obvious craft sits one level up: sometimes a statement P is unprovable by induction while a *stronger* statement sails through — because the inductive step gets to *assume* the hypothesis, and a stronger claim is also a stronger tool. Pólya named the effect the inventor's paradox: proving more can be easier. It is the single most counterintuitive lever in the beginner-to-olympian transition.

The smell. Any claim parameterized by n ; recursive or self-similar structures — trees, tilings of 2^n boards, nested expressions. The strengthening trigger is specific and worth memorizing: your inductive step *almost* works, but the hypothesis is too weak to push it through — it needs a fact the plain statement does not carry. The response is not to abandon induction. It is to load the statement with the missing cargo — a sharper bound, a more general configuration — and re-run.

★★ • The Sum of Odd Numbers

Prove that $1 + 3 + 5 + \cdots + (2n - 1) = n^2$.

Pause. Draw it before you prove it.

Solution. Base, $n = 1$: $1 = 1^2$. Step: assume the sum to $(2n - 1)$ is n^2 ; add the next odd number: $n^2 + (2n + 1) = (n + 1)^2$. ■ The picture is better than the

algebra: an $n \times n$ square grows into an $(n + 1) \times (n + 1)$ square by wrapping an L-shaped shell of exactly $2n + 1$ cells around two sides. Induction as the geometry of accretion.

The Move — Climb from n to $n + 1$. The plainest tool in the book — and the chassis every refinement below bolts onto.

★★★ • The Broken Board

Prove that every $2^n \times 2^n$ board with any one cell missing can be tiled by L-trominoes (three cells in an L).

Pause here. Chapter 9 proved boards impossible; now build one.

Solution. The strengthening is already in the statement — and that is the lesson. The "natural" question (a corner missing) is harder to induct on: the quartered board hands you sub-boards with no missing corners, and the hypothesis has no grip. Prove the stronger "any cell" version instead. Base, $n = 1$: a 2×2 board minus any cell is an L-tromino. Step: split the 2^{n+1} board into four 2^n quadrants; the missing cell lies in one of them. Place a single tromino at the center, covering the corner cell of each of the *other three* quadrants. Now every quadrant is a $2^n \times 2^n$ board missing exactly one cell — apply the stronger hypothesis four times. ■ (Count check: $(4^{n+1} - 1)/3 = 4 \cdot (4^n - 1)/3 + 1$ — three quadrant-deficiencies manufactured, one tromino spent. ✓)

This is the constructive twin of Chapter 9's colorings: pigment proves impossibility, induction manufactures possibility.

The Move — Generalize until the step has a grip. A more general claim hands the inductive step more to hold onto. If quartering the problem produces sub-problems your hypothesis cannot touch, widen the hypothesis until it can.

★★★★ • A Cushion Made of Nothing

Prove that $1/1^2 + 1/2^2 + \dots + 1/n^2 < 2$ for every n .

Pause. Try the obvious induction and watch it die.

Solution. The obvious induction: assume $S_n < 2$; then $S_{n+1} = S_n + 1/(n+1)^2 < 2 + \text{something}$. No room. The statement is true but inductively sterile — the hypothesis leaves nothing to absorb the new term.

Strengthen it: prove $S_n \leq 2 - 1/n$. Base, $n = 1$: $S_1 = 1 = 2 - 1$. Step: given $S_n \leq 2 - 1/n$, it suffices that $1/(n+1)^2 \leq 1/n - 1/(n+1) = 1/(n(n+1))$, i.e. that $n(n+1) \leq (n+1)^2$, i.e. that $n \leq n+1$. True. So $S_{n+1} \leq 2 - 1/(n+1)$, and in particular $S_n < 2$ for all n . ■

The stronger claim carries its own fuel: the $-1/n$ term is exactly the absorbing cushion the step needed. (Euler later showed the true limit is $\pi^2/6 \approx 1.645$ — our cushion is nowhere near tight, and that is fine. A strengthening need only be *inductively closed*, not sharp.)

The Move — Strengthen the hypothesis. An induction hypothesis is not what you must prove; it is what you get to use. When the step leaks, ask Pólya's inverted question: what stronger statement would make this step trivial?

Machines and this move. Hypothesis strengthening is lemma invention — the classic bottleneck of automated theorem proving, with heuristics but no reliable method — and models mimic famous strengthenings well while rarely originating the right generalization fresh; humans feel *where the step leaks* and patch the statement.

Recursion → Dynamic Programming

The tool. Express the answer to a problem in terms of answers to smaller instances; when subproblems repeat, cache them (memoization) or fill them bottom-up (dynamic programming). Textbooks sell DP as a technique. Competitors know it as a *representation question*: choose what a "state" is — the minimal summary of the past that the future needs — and the recur-

rence usually writes itself. The classical ladder climbs one dimension at a time: Fibonacci (one dimension — the ways to tile a $1 \times n$ strip with 1s and 2s), grid paths (two), knapsack (a *resource*), bitmask DP (a set), digit and tree DP (structured prefixes).

The smell. "Count the ways...", "minimum cost to...", "longest ... over sequences, grids, trees." The problem decomposes by "what was the last choice?" Brute force is exponential, but the *relevant* past is small. The professional's checklist, in order: What are the decisions? After a prefix of decisions, what does the future actually need to know? — that is the state. Transitions are one decision; the cost is $\#states \times$ transition cost; and if that is too big, the state is carrying junk — compress it.

★★ • Twenty Paths

How many monotone paths — steps right and down only — run from the top-left to the bottom-right corner of a 4×4 lattice of nodes, from (0,0) to (3,3)?

Pause. Solve the top-left corner of the question first.

Solution. State: $f(i, j)$ = the number of monotone paths reaching node (i, j) . You arrived from above or from the left, so $f(i, j) = f(i-1, j) + f(i, j-1)$, with the edges all 1. Fill the table:

	j=0	1	2	3
i=0	1	1	1	1
1	1	2	3	4
2	1	3	6	10
3	1	4	10	20

Answer: **20**. And look at the table — it is Pascal's triangle, tilted: $f(i, j) = C(i+j, i)$, here $C(6,3) = 20$. ✓ ■

The Move — Name the state, fill the table. The recursion ("the last step came from left or above") is trivial; the payoff is a closed form materializing out of a table. Dynamic programming as experimental mathematics.

★★★★ • The Knapsack, End to End

Capacity $W = 8$. Items: A (weight 2, value 3), B (3, 4), C (4, 5), D (5, 8). Maximize total value within the capacity — and know *why* the greedy fails.

Pause here. Try greed first, deliberately.

Solution. Greedy by value density: $D = 1.6$, $A = 1.5$, $B \approx 1.33$, $C = 1.25$. Greed takes D (weight 5), then A (weight 2) — value 11, weight 7, nothing else fits. **11 is not optimal.**

Now the DP. State: $dp[c]$ = the best value achievable within capacity c , over the items processed so far; process items one by one, scanning c downward so each item is used at most once. Start all-zero.

- after A: $dp = [0, 0, 3, 3, 3, 3, 3, 3]$
- after B: $dp = [0, 0, 3, 4, 4, 7, 7, 7]$ (e.g. $dp[5] = \max(3, dp[2] + 4) = 7$)
- after C: $dp = [0, 0, 3, 4, 5, 7, 8, 9]$ ($dp[6] = \max(7, dp[2] + 5) = 8$; $dp[7] = \max(7, dp[3] + 5) = 9$)
- after D: $dp = [0, 0, 3, 4, 5, 8, 8, 11, 12]$ ($dp[7] = \max(9, dp[2] + 8) = 11$; $dp[8] = \max(9, dp[3] + 8) = 12$)

Optimum **12 = B + D**: weights $3 + 5 = 8$, values $4 + 8$ — the greedy's 11 loses to 12. (Verified against full enumeration of all 2^4 subsets: the feasible bests are $A+D = 11$, $B+D = 12$, $B+C = 9$, $A+C = 8$.) Cost: 4 items $\times$ 9 capacities ≈ 36 cell-updates replace 16 subset checks — and at $n = 100$, $W = 100,000$, the same table replaces 2^{100} subsets. One honest caveat: $O(nW)$ is *pseudo-polynomial* — W is exponential in its own digit count — so knapsack remains NP-hard.



The Move — Compress history into a sufficient statistic. The state is not "which items were taken" (2^n states) but "how much capacity is spent" — because value is all the future needs and capacity is all the future is constrained by. Every hard DP is hard exactly here: longest-increasing-subsequence compresses "the whole prefix" to "smallest tail per length"; traveling-salesman DP compresses "the route so far" to "(visited set, current city)"; digit DP compresses "the number built" to "(position, still tight?, property so far)." The same move, up the ladder.

Machines and this move. The most machine-friendly tool in the arsenal — once states are defined, filling tables is what computers *are*, and the ladder saturates their training data; the residual human edge is state *invention* — spotting that an exotic history collapses to a small sufficient statistic.

Greedy, with the Exchange Argument

The tool. A greedy algorithm makes the locally best choice and never revisits it. Greed is usually wrong; when it is right, the standard proof is the **exchange argument**: take any optimal solution OPT, show you can swap one of its choices for the greedy's choice without making it worse, and induct — so the greedy's output matches OPT. (The deep-structure version says greedy is exactly optimal on matroids; competitors mostly wield the exchange lemma directly.)

The smell. Scheduling, selecting, merging, ordering — anywhere a sort *by the right key* feels plausible, and the future depends on the past only through a running total or frontier. The disciplined reflex is double-sided: hunt for the exchange *and* hunt for a counterexample with equal energy — small n , adversarial weights. If neither wins in ten minutes, suspect DP instead. And carry the warning shot every competitor eats early: coins $\{1, 3, 4\}$, amount 6. Greedy: $4 + 1 + 1$, three coins. Optimal: $3 + 3$, two. Greedy-with-proof or bust.

★★ • Picking Meetings

From the intervals $[0,2)$, $[1,3)$, $[2,4)$, $[3,5)$, $[4,6)$, choose the maximum number that pairwise do not overlap.

Pause. Which key would you sort by?

Solution. The greedy rule: repeatedly take the interval with the **earliest finish** among those compatible with what you have. Trace: take $[0,2)$; skip $[1,3)$; take $[2,4)$; skip $[3,5)$; take $[4,6)$ — **3 meetings**, and 3 is the maximum (four pairwise-disjoint intervals of length 2 would need total length 8 in a span of 6).

The exchange proof of the rule: let g_1 be the earliest-finishing interval overall and o_1 the earliest element of any optimal set OPT. Then $\text{finish}(g_1) \leq \text{finish}(o_1)$, so $\text{OPT} - \{o_1\} + \{g_1\}$ is still pairwise-disjoint — g_1 ends no later than o_1 , so it cannot collide with the rest of OPT — and has the same size. Some optimum therefore starts with the greedy's choice; recurse on the remaining span. ■

The Move — Morph the optimum into the greedy. The proof never says greedy beats OPT. It says OPT can be reshaped into the greedy, one lossless swap at a time — a template reusable across half of algorithmics.

★★★★ • The Largest Number

Given positive integers — say $\{9, 5, 34, 3, 30\}$ — arrange them in a row so that the concatenation of their decimal digits is the largest possible number.

Pause here. The obvious sort is wrong; find the right comparator.

Solution. Sort by the comparator: **x before y when the concatenation xy exceeds yx** (compared as numbers). Trace: $9 \text{ vs } 5 \rightarrow "95" > "59"$; $34 \text{ vs } 3 \rightarrow "343" > "334"$, so 34 precedes 3; $3 \text{ vs } 30 \rightarrow "330" > "303"$, so 3 precedes 30.

Order: 9, 5, 34, 3, 30 — answer **9534330** (beating, for instance, 9, 5, 3, 34, 30 → 9533430).

Why this is legitimate — two obligations most solvers skip:

1. **The comparator is a genuine total order.** Writing X for x 's value and ℓ_x for its digit-length: $xy > yx \iff X \cdot 10^{\ell_y} + Y > Y \cdot 10^{\ell_x} + X \iff X(10^{\ell_y} - 1) > Y(10^{\ell_x} - 1) \iff X/(10^{\ell_x} - 1) > Y/(10^{\ell_y} - 1)$. The funny pairwise rule is secretly a sort by the single real key $\kappa(x) = X/(10^{\ell_x} - 1)$ — hence transitive, hence a valid sort. And κ has a lovely reading: $\kappa(34) = 34/99 = 0.343434\dots$, $\kappa(3) = 3/9 = 0.333\dots$ — each number's digits repeated forever. The arrangement sorts by *what you would become if you repeated without end*.
2. **Exchange for optimality.** In any arrangement, if some *adjacent* pair sits against the comparator (...yx... where $xy > yx$), swapping them changes only that junction — every other digit keeps its position — and strictly increases the whole number. Any non-sorted arrangement contains such an adjacent inversion, so it is not optimal; the sorted arrangement is the maximum, with κ -ties interchangeable. ■

The Move — Prove both duties. A greedy rule owes two debts: *coherence* (the comparator must be transitive — and a programming language will happily "sort" with a non-transitive one into nonsense) and *exchange-stability*. Pay both or don't submit.

Machines and this move. Language models propose greedy rules fluently and hallucinate their correctness; the {1, 3, 4} reflex — attacking your own rule before trusting it — is a trained adversarial habit, not a default.

Binary Search on the Answer

The tool. To optimize a quantity that is hard to compute directly, flip the question: *is there a solution with value at least x ?* If feasibility is **monotone** in x — feasible at x implies feasible at everything on one side — then the optimum is a threshold, and binary search finds it with $O(\log \text{range})$ calls to a *decision* procedure. You have traded "optimize" for "decide," and deciding is almost always easier: a greedy or a graph check.

The smell. The magic words: *minimize the maximum, maximize the minimum* — "largest minimum distance," "smallest time by which all jobs finish." Also: the answer is a single number in a bounded range, and checking a candid-

ate is easy even though producing the best one is not. The reflex: ask "*if an oracle whispered the answer x , could I verify a solution at level x cheaply?*" If yes, and monotone — the problem is solved; all that remains is the feasibility subroutine.

★★ • The Sawmill

Trees of heights 20, 15, 10, 17. A blade set at height H cuts off everything above H , collecting $\sum \max(0, h_i - H)$ units of wood. Find the largest integer H that collects at least 7 units. (A classic from the training sites, known there as EKO.)

Pause. Don't optimize. Guess and check — efficiently.

Solution. Feasibility $f(H) = \text{"wood}(H) \geq 7$ ". Wood is non-increasing in H , so f is monotone: true low, false high. Binary search on $[0, 20]$:

- $H = 10$: wood = $10 + 5 + 0 + 7 = 22 \geq 7$ ✓ $\rightarrow$ go higher, $[10, 20]$
- $H = 15$: wood = $5 + 0 + 0 + 2 = 7 \geq 7$ ✓ $\rightarrow [15, 20]$
- $H = 18$: wood = $2 + 0 + 0 + 0 = 2$ ✗ $\rightarrow [15, 17]$
- $H = 16$: wood = $4 + 0 + 0 + 1 = 5$ ✗ $\rightarrow [15, 15]$

Answer **$H = 15$** (wood(15) = 7 ✓, wood(16) = 5 ✗). Five decision calls replace a scan of all heights; with heights up to 10^9 it is about thirty calls. The logarithm is the whole business model. ■

The Move — Interrogate candidates, not the optimum. If a whispered answer would be easy to check, stop trying to compute it and start bisecting for it.

★★★★ · Aggressive Cows

Stalls at positions 1, 2, 4, 8, 9 on a line. Place 3 cows in stalls so that the minimum distance between any two is as large as possible. (The classic is known as AGGRCOW.)

Pause here. Two tools will have to compose.

Solution. Decision version: *can 3 cows be placed with every pair at least d apart?* Check greedily: put the first cow in the leftmost stall, then always take the next stall at least d beyond the last cow; count placements. The check needs its own small exchange proof: if any valid placement of k cows at gap d exists, sliding its leftmost cow to the leftmost stall and each next cow to the earliest stall $\geq d$ beyond the previous only moves cows leftward or not at all, never violating d — so the leftmost-greedy placement seats at least as many cows. Hence "greedy seats ≥ 3 " is equivalent to feasibility. Feasibility is monotone — bigger d is harder — so binary search d on $[1, 8]$:

- $d = 4$: place at 1 $\rightarrow$ next $\geq 5 \rightarrow 8 \rightarrow$ next $\geq 12 \rightarrow$ none. Two cows $< 3 \times \rightarrow [1, 3]$
- $d = 2$: 1 $\rightarrow 4 \rightarrow 8$. Three cows $\checkmark \rightarrow [2, 3]$
- $d = 3$: 1 $\rightarrow 4 \rightarrow 8$, gaps 3 and 4. Three cows $\checkmark \rightarrow [3, 3]$

Answer **$d = 3$** , with $d = 4$ shown infeasible. ■

The ★★★★★ content is the frame: the solver never optimized anything. "Maximize the minimum" dissolved into (a) spotting monotonicity, (b) a greedy decision check carrying its own exchange proof, (c) logarithmically many repetitions. Tools composing tools — this section outside, the previous one inside.

The Move — Dissolve "maximize the minimum" into feasibility. The optimum is a threshold; hunt the threshold, not the solution.

Machines and this move. Fully mechanical once monotonicity is asserted — this may be the single most common pattern in models' competitive-programming training data; the human step is certifying monotonicity on

messy problems, where "obviously monotone" occasionally isn't. Prove the one-line implication before trusting the search.

Working Backwards

The tool. Start from the goal state and compute predecessors, instead of exploring forward from the start. In games this is backward induction: label terminal positions Won or Lost, then propagate — a position is Winning if some move reaches a position Lost for the opponent, Losing if all moves reach Winning ones. In puzzles it is "what must the last step have been?" — the last step is often far more constrained than the first, so the tree is thinner from that end.

The smell. A clearly specified goal with a foggy start: "reach exactly 100," "end with one token," "mate in three." Games of perfect information — "who wins with best play?" Moves easier to invert than to choose. The reflex: when forward branching is wide and the goal is narrow, grow the tree from the narrow end — or from both ends at once, halving the exponent.

★★ • The Race to 100

Two classics, one skeleton. (a) 21 stones; players alternately remove 1, 2, or 3; whoever takes the last stone wins. (b) Players alternately add 1 through 10 to a running total starting at 0; whoever says 100 wins. Who wins each, and how?

Pause. Start from the end — literally.

Solution. (a) Label backwards. Facing 0 stones, you have lost — the previous player took the last. Facing 1–3, you win (take them all). Facing 4, every move hands the opponent a win: 4 is Losing. Propagating, the Losing positions are exactly the multiples of 4 — from $4k$ any take of t leaves a non-multiple; from $4k + r$ ($r = 1, 2, 3$), take r . Since $21 \equiv 1 \pmod{4}$, the first player wins: take 1, then answer every t with $4 - t$, restoring a multiple of 4. (b) The same engine, additive: the targets to *hand your opponent* run backwards from 100 in steps of 11 — 100, 89, 78, ..., 12, 1 — since your k is answered by $11 - k$. The first player says **1**, then climbs 12, 23, ..., 89, 100. ■

Forward play at 21 sees a bushy tree; backward labeling yields the entire theory as one arithmetic progression. It also plants the general concept — cold positions (Losing for the mover) versus hot — the periodic skeleton under Nim-like game theory: Chapter 9's balanced columns were exactly Nim's cold positions, and the general theory (Sprague–Grundy) grows from this seed.

The Move — Label from the end. The goal is one state; the start is many. Compute Won/Lost outward from the goal and read the answer off your own table.

★★★★ • Wythoff's Game, and the Golden Ratio

Two piles of stones. A move: take any positive number from one pile, **or** take the *same* positive number from both. Last stone wins. Characterize the losing positions.

Pause here. Grind small cases before you theorize.

Solution. Backward-label small states, pairs unordered. (0,0) is Losing — the mover faces no stones, the previous player took the last. Everything that reaches (0,0) in one move — every (k,0) and (k,k) — is Winning. Scan upward for the next all-exits-Winning state: (1,2). Its moves reach (0,2), (1,1), (1,0), and diagonally (0,1) — each Winning (each reaches (0,0) in one). So (1,2) is Losing. Prune everything that can reach (0,0) or (1,2), and the next candidate is (3,5). Its single-pile moves reach (0,5), (1,5), (2,5), (3,4), (3,3), (3,2), (3,1), (3,0); its diagonal moves reach (2,4), (1,3), (0,2). Every one is Winning: (0,5), (3,0), (0,2) hit (0,0) directly and (3,3) does diagonally; (1,5) → (1,2) by taking 3; (2,5) → (2,1) by taking 4; (3,2) → (1,2) by taking 2; (1,3) → (1,2) by taking 1; (3,1) → (2,1) by taking 1; (3,4) → (1,2) by the diagonal -2; (2,4) → (2,1) by taking 3. (Pairs are unordered: (2,1) is (1,2).) So (3,5) is Losing. The sieve continues: (0,0), (1,2), (3,5), (4,7), (6,10), (8,13), ...

Two structural facts fall out of the table: every positive integer appears in exactly one Losing pair, and the k-th pair differs by exactly k. Closed form (Wythoff, 1907): the k-th losing position is **($\lfloor k\varphi \rfloor$, $\lfloor k\varphi^2 \rfloor$)**, with $\varphi = (1 + \sqrt{5})/2$

2. Check — $k = 1$: (1, 2); $k = 2$: ([3.236], [5.236]) = (3, 5); $k = 3$: ([4.854], [7.854]) = (4, 7) ✓. Why the golden ratio: Beatty's theorem says the sequences $[k\alpha]$ and $[k\beta]$ partition the positive integers exactly when $1/\alpha + 1/\beta = 1$ — and $1/\varphi + 1/\varphi^2 = 1$; while $\varphi^2 = \varphi + 1$ makes the k -th gap exactly k . Those two properties are precisely the sieve's rules: each new Losing pair uses the smallest unused integer (the partition property) with gap k (the diagonal-move property). ■

Retrograde analysis was the experiment; the golden ratio is the theory. The honest workflow — grind the backward table, stare, conjecture, then prove via Beatty — is exactly how the result was found, and it models a two-stroke engine this book keeps meeting: machinery generates data; recognition compresses it; proof certifies it.

The Move — Grind the table, then earn the formula. Backward induction gives you truth without insight; compressing the table into a closed form is the insight — and it must then be proved, not pattern-matched.

Machines and this move. Retrograde analysis is the machine's home turf — chess endgame tablebases, every position of seven pieces or fewer solved backward, are its industrial monument; the human stroke is compressing the table into φ , which no tablebase will ever volunteer.

The five, at a glance

Move	The smell	Emblem
Strengthen the hypothesis	the inductive step almost works	the $2 - 1/n$ cushion
Compress history into a state	count/optimize over sequences; small relevant past	knapsack's 12 beats greed's 11
Exchange argument	a sort-by-key feels plausible	9534330
Binary search the answer	minimize-the-maximum	cows at $d = 3$

Move	The smell	Emblem
Label from the end	narrow goal, foggy start	Wythoff's ϕ

Builders, all five: each returns not just an answer but a construction with a warranty. One family remains — the tools that change what a problem is before anyone starts solving it.

CHAPTER 11

The Modeler's Eye

The last family does not attack a problem; it changes what the problem is. Re-representation — story into graph, sum into set, puzzle into average — is the quietest move in the arsenal and the most powerful: after the right translation, famously hard problems finish themselves. It is also where human solvers hold their widest lead over machines. The chapter ends with three famous set pieces, told with the record straightened.

Story → Structure

The tool. Recast the problem as an instance of a known, solved abstraction: a graph, a flow network, a matching, a shortest path, a system of congruences. The solving act is a translation — say what the vertices are, what the edges are, and which classical theorem now applies. Competitive programming at the top level is a reduction-recognition sport; most hard problems are a known kernel wearing a costume.

The smell. Entities with pairwise relationships: a graph is coming. "Assign X to Y" under compatibility constraints: bipartite matching. "Throughput," "bottleneck," "cut them off": max-flow, min-cut. Sequential states with legal transitions: shortest path over a *state space*, not a physical space. The meta-signal: the story resists direct attack, yet every sentence sounds like it is describing a structure. The reflex: **name the vertex**.

The book's first puzzle was secretly this section's easy example. Chapter 1's wolf, goat, and cabbage becomes: vertex = a *legal* distribution of {farmer, wolf, goat, cabbage} over two banks — 16 raw states, 10 legal; edge = one crossing. The question is now a shortest path; breadth-first search finds the 7-crossing route, *certifies* it minimal — and the famous aha, bringing the goat back, becomes just an edge like any other, crossed by the same dumb search. The insight was in the modeling; after it, no insight is required. The

same move solves Rubik's cube, sliding puzzles, and protocol verification: shortest paths in graphs of *configurations*, not places.

★★★★ • Guards in the Palace

A 4×4 palace floor has treasures at cells (1,1), (1,3), (2,2), (3,1), (3,3) — row, column. A guard patrols one full row or one full column. Find the minimum number of guards covering all treasures, and *prove* it minimal.

Pause here. What, precisely, is a vertex?

Solution. Go bipartite: left vertices are rows, right vertices are columns, and **each treasure is an edge** joining its row to its column. A guard is a *vertex*; covering all treasures means touching every edge: a **vertex cover**. In bipartite graphs, minimum vertex cover equals maximum matching — König's theorem, 1931. Lower bound: treasures (1,1), (2,2), (3,3) form a matching of size 3, sharing no row and no column, and any guard covers at most one edge of a matching — so 3 guards are forced. Upper bound: rows 1, 2, 3 cover all five treasures. Minimum: **3**. ■ (Direct check that two fail: covering (2,2) takes row 2 or column 2, which touches none of the four corner treasures — and one line covers at most two of those four.)

Nothing in the story says "graph," and the winning translation — treasure = *edge*, not vertex — is exactly what modeling skill consists of. Once made, a 1931 theorem finishes the fight, certificate included: the matching is the lower-bound witness.

The Move — Name the vertex. Get the atoms right and the theorem is usually waiting.

Machines and this move. Executing a max-flow is trivial; *deciding the story is a flow problem* is abduction over an open representation space — models recognize costumes they have read and fail on fresh disguises.

Casework & WLOG

The tool. When no single argument covers everything, partition the space into finitely many cases and kill each. The discipline: exhaustive (say why), preferably disjoint, and *few* — which is where "without loss of generality" earns its keep: quotient out symmetry *before* splitting, by fixing an ordering, a position, a naming.

The smell. A small structured space — two colors, residues mod 3, who moves first. A statement invariant under relabeling: that invariance is a coupon; spend it. Warning smell in the other direction: a case count past five or so usually means a missing invariant or extremal choice. Casework is the honorable last resort.

★★ • The Party of Six

Prove that among any 6 people, some 3 mutually know each other or some 3 are mutual strangers.

Pause. Fix one guest and look at their five relationships.

Solution. Fix a person v . Of the 5 relations at v , some kind appears at least $\lceil 5/2 \rceil = 3$ times (pigeonhole); **WLOG it is "know"** — the statement is symmetric under swapping know/stranger, so this halves the proof. Say v knows A, B, C . If some pair of them know each other, that pair plus v is a knowing triangle. If no pair does, A, B, C are three mutual strangers. Done either way. ■ Sharpness: seat 5 people in a pentagon, neighbors acquainted, diagonals strangers — no triangle of either kind. So 6 is exact: $R(3,3) = 6$.

Three tools in eight lines — pigeonhole picks the branch, WLOG halves the tree, excluded middle makes the cases exhaustive. Ramsey theory's slogan follows: structure becomes unavoidable at scale.

The Move — Quotient the symmetry before you split. Every symmetry you fail to spend becomes two identical cases written out twice.

★★★★ • Three of Five

Prove that among any 5 integers, some 3 have a sum divisible by 3 — and show that 4 integers do not suffice.

Pause here. Choose the case split so each branch has a weapon.

Solution. Only residues mod 3 matter; sort the five into classes R_0, R_1, R_2 and split on the *maximum class size*. **Case A — some class holds ≥ 3 .** Take three from it: $\text{sum} \equiv 3r \equiv 0 \pmod{3}$. **Case B — every class holds ≤ 2 .** Five numbers then force sizes (2, 2, 1) in some order — in particular no class is empty, since two classes hold at most 4. One number from each class: $0 + 1 + 2 \equiv 0 \pmod{3}$. The cases are exhaustive: the maximum is either ≥ 3 or ≤ 2 .

■ **Sharpness.** Four integers dodge: residues $\{0, 0, 1, 1\}$ — any three sum to 1 or 2 (mod 3), never 0. So 5 is exact. This is the $n = 3$ case of Erdős–Ginzburg–Ziv (1961): among any $2n - 1$ integers, some n sum divisibly by n , and $2n - 2$ never suffice — $n - 1$ zeros and $n - 1$ ones being the general dodge. Chapter 8's "strange exactness" signpost, detonating on schedule.

The Move — Design cases backward from the finishing moves. Split on a variable chosen so every branch dies to an available weapon; an analysis that cannot exhibit tightness is often hiding a slack case.

Machines and this move. Exhaustive checking is the machine's native gait — Appel and Haken's four-color proof (1976) checked about 1,936 configurations — but the WLOG license, seeing which quotient collapses 3^6 cases to 2, remains a human design act.

Solve Smaller First

The tool. Two directions of one dial. *Small cases first:* compute $n = 1, 2, 3, 4$ by hand, conjecture the pattern, then **prove** it — the computed cases become the induction's base. *Generalize first* — Pólya's inventor's paradox again: replace the specific question (100, this year's number, this board) with its parameterized family; the general question has more structure to grab, and the specific answer falls out by substitution.

The smell. A specific large constant — 2026, 100, 10^{18} — is almost always a costume over a clean parameter n . "Find the value" with no visible structure: tabulate immediately. A statement that feels arbitrary: generalize until it feels inevitable. And the professional's reflex: a pattern from three data points is a **hypothesis, not an answer**.

★★ • The Last Digit of 7^{2026}

Find the last digit of 7^{2026} .

Pause. Replace 2026 with n and make a small table.

Solution. Tabulate $7^n \bmod 10$: 7, 9, 3, 1, then 7, 9, 3, 1, ... — period 4. (Once a residue repeats, the whole cycle must: multiplication by 7 mod 10 is a fixed map.) So the answer depends only on $n \bmod 4$, and $2026 \equiv 2 \pmod{4}$: last digit = $7^2 \bmod 10 = 9$. ■ Behind the toy, a general machine: powers mod m are eventually periodic — pigeonhole on m residues — and Euler's theorem predicts the period divides $\phi(10) = 4$.

The Move — Replace the year with n . A two-line periodicity in a four-digit costume. Undress it.

★★★★ • Circles Against the Siren

n circles in the plane, each pair meeting in exactly 2 points, no three through a common point. Into how many regions do they cut the plane? Compute small cases, conjecture — then prove.

Pause here, and do not trust your first formula.

Solution. Small cases: $n = 1 \rightarrow 2$ regions; $n = 2 \rightarrow 4$; $n = 3 \rightarrow 8$. The siren sings: 2^n . **Test $n = 4$.** The fourth circle crosses the 3 others in 6 points, is cut into 6 arcs, and each arc splits one region: $8 + 6 = 14$, **not 16**. The doubling was a mirage of small numbers. The incremental count, now honest: the k -th circle meets its $k - 1$ predecessors in $2(k - 1)$ points, is cut into $2(k - 1)$ arcs, and each arc splits exactly one existing region — adding $2(k - 1)$, while the

first circle adds 1 to the initial 1. $R(n) = 2 + \sum_{k=2}^n 2(k-1) = n^2 - n + 2$. Check: 2, 4, 8, 14, 22 ✓. Then certify *independently* with Euler's formula $V - E + F = 2$: $V = n(n-1)$ intersection points; each circle carries $2(n-1)$ of them, so $E = 2n(n-1)$ arcs; $F = 2 - V + E = n(n-1) + 2$ — the same $n^2 - n + 2$. ■ Two roads, one answer: the actual standard of "verified."

Keep the cautionary twin beside it — Moser's circle problem: n points on a circle, all chords drawn, no three concurrent; count regions. The sequence runs 1, 2, 4, 8, 16 — then **31** at $n = 6$. The true formula is $C(n,4) + C(n,2) + 1$ (each chord adds a region; each interior crossing, one per 4 points, adds one more): 1, 2, 4, 8, 16, 31, 57 — A000127 in the OEIS. Five consecutive doublings, then betrayal. Small cases propose; only proofs dispose.

The Move — Conjecture cheaply, verify expensively, certify twice. The fourth small case is cheap insurance; a second independent derivation is the seal.

Machines and this move. Machines out-enumerate us absurdly — competitors really do compute four terms and consult the OEIS — but extrapolating 2^n from four terms is Moser's trap industrialized; conjecture-then-certify transfers verbatim to auditing machine output.

Counting Twice

The tool. Count one set of objects by two different bookkeeping schemes; both count the same thing, so the two expressions are equal — and that equation is the theorem. The usual object is a set of *incidences*: the row sums versus the column sums of one 0/1 matrix. Double counting is combinatorics' identity-forgery.

The smell. An identity or divisibility between count-like quantities; a sum that would simplify reindexed — "for each student..." versus "for each problem..."; handshakes, memberships, averages (an average is a double count divided by a headcount). The reflex: define the incidence set explicitly — "pairs (x, y) such that..." — then total it both ways.

★★ • Handshakes, and a Census

(a) Prove that in any graph the degrees sum to twice the number of edges, and conclude that the number of odd-degree vertices is always even. (b) In a class, 35 students each solved exactly 3 problems, and every problem was solved by exactly 5 students. How many problems?

Pause. Both are the same move.

Solution. (a) Count the incidences (v, e) with vertex v an endpoint of edge e : grouped by vertex, $\sum \deg(v)$; grouped by edge, exactly 2 each. So $\sum \deg(v) = 2|E|$ — and a sum of integers that is even must contain an even number of odd terms, so odd-degree vertices come in even supply. (b) Count (student, problem) incidences twice: $35 \cdot 3 = P \cdot 5$, so $P = 21$. ■ No enumeration, no structure — two viewpoints alone pin the answer.

The Move — Total the matrix both ways. Rows or columns, the grand total cannot care. Make that indifference say something.

★★★★ • Every Fixed Point, Counted Sideways

(IMO 1987, Problem 1.) Let $p_n(k)$ be the number of permutations of $\{1, \dots, n\}$ with exactly k fixed points. Prove that $\sum_{k=0}^n k \cdot p_n(k) = n!$.

Pause here. What single set might both sides be counting?

Solution. The left side counts each permutation π once *per fixed point* it has: it is the size of the incidence set $F = \{(\pi, i) : \pi(i) = i\}$, grouped by π . Group by i instead: the permutations fixing a given i are exactly the permutations of the other $n - 1$ elements — $(n - 1)!$ of them. So $|F| = n \cdot (n - 1)! = n!$. ■ Sanity at $n = 3$: one permutation has 3 fixed points, three have 1, two have 0; $3 \cdot 1 + 1 \cdot 3 + 0 \cdot 2 = 6 = 3!$ ✓.

An IMO problem — and the entire solution is one change of viewpoint, three lines long. Difficulty and length-of-best-solution are nearly independent variables; the exam tested whether that sum *smells like* an incidence count.

Read probabilistically, the identity says the average permutation has exactly one fixed point — which hands us the next tool.

The Move — Every sum is a set being counted; find the set.

Machines and this move. Verifying that two counts agree is arithmetic; *choosing the incidence set* is a re-representation act that generalizes poorly in models on novel weighted sums.

The Probabilistic Lens

The tool. Import probability into problems that mention none. Engine one — **linearity of expectation**: $E[X_1 + \dots + X_n] = \sum E[X_i]$ with *zero* independence assumptions, so tangled, correlated messes decompose into per-part sums. Engine two — **averaging**, Erdős's probabilistic method: if the average over a family is v , some member achieves at least v . Pick a random object, compute its expected merit, and a good object exists — never constructed.

The smell. "Prove there exists an arrangement / coloring / subset with at least..." and no visible construction: randomize and average. Any quantity that is a sum of many small overlapping contributions: indicator sums. The reflex: **name the indicators** — and watch dependence, the great obstacle everywhere else, simply not matter.

★★ • The Average Permutation

Show that the average number of fixed points of a random permutation of n elements is 1 — for every n .

Pause. No sums over k allowed.

Solution. Let $X_i = 1$ if $\pi(i) = i$, else 0; the number of fixed points is $X = \sum X_i$, and $E[X_i] = P(\pi(i) = i) = (n-1)!/n! = 1/n$. Linearity — the X_i are heavily correlated, and it does not matter: $E[X] = n \cdot (1/n) = 1$. ■

The arithmetic is the IMO identity from one section back, divided by $n!$. But the double count needed an inspiration — find the incidence set — while the expectation needed only a standing reflex. Probabilistic language converts

one-off insights into routine calculations; that conversion is the point of owning a lens.

The Move — Write the indicators. Linearity never asks whether they are independent.

★★★★ • The Big Cut

Prove that every graph with m edges has a partition of its vertices into two sides with at least $m/2$ edges crossing between them. (No randomness in the statement.)

Pause here. Flip a coin for every vertex.

Solution. Color each vertex Left or Right by independent fair coin. For an edge $e = \{u, v\}$, let X_e indicate "u and v on opposite sides": $P(X_e = 1) = 1/2$, so $E[\sum X_e] = m/2$. An average of $m/2$ forces some outcome $\geq m/2$ — an average cannot exceed its maximum. That outcome is the partition. ■

Then derandomize, with an old friend. The proof names no cut, but a monovariant (Chapter 9) extracts one: start anywhere; while some vertex has more neighbors on its own side than across, flip it. Each flip strictly increases the crossing count, which is bounded by m , so the process halts — and at a halt every vertex has at least half its degree crossing, so crossings $\geq \frac{1}{2} \cdot \sum \deg(v)/2 = m/2$. Local search finds what probability promised. (One rung up: a random assignment satisfies $7/8$ of any 3-SAT formula's clauses, each failing with probability $1/8$ — and Håstad proved $7/8$ optimal to approximate unless $P = NP$. The trivial random bound is the exact frontier of tractability.)

Erdős founded the method in 1947, lower-bounding Ramsey numbers with random colorings. It is average-case thinking as a weapon: ask what a typical case achieves, then upgrade typicality to existence (averaging) or to an algorithm (derandomization). Chapter 7's hundred-prisoner miracle is this lens at full power — no individual's odds improve; the strategy merely herds

the failures into the same unlucky worlds. The mean is untouchable; the variance is the lever.

The Move — Upgrade the average to existence, then to an algorithm.

Machines and this move. There is nothing for a search to find until someone reframes "exists" as "average" — owning the decision to randomize a deterministic world is still the human part.

Budget First

The tool. Before solving, budget: the size of the search space, the information content of the answer, the operations affordable — and let those numbers *choose the method*. For programmers it is literal: about 10^8 simple operations per second, so the constraints encode the intended algorithm. For mathematicians it is information-theoretic: a procedure with k ternary outcomes distinguishes at most 3^k situations, so any strategy needs $\lceil \log(\#cases) \rceil$ steps — a lower bound requiring no cleverness, only counting.

The competitive programmer's decoder ring, worth printing:

constraint	affordable shape	typical intent
$n \leq 11$	$n! \approx 4 \cdot 10^7$	brute-force permutations
$n \leq 20-24$	$2^n \cdot n$	bitmask DP
$n \leq 40$	$2^{n(n/2)}$	meet in the middle
$n \leq 500$	n^3	cubic DP / all-pairs paths
$n \leq 10^5-10^6$	$n \log n$	sort / heap / divide & conquer
$n \leq 10^8$	n	linear scan, two pointers
$n \sim 10^{18}$	$\log n / \sqrt{n}$	binary search on the answer, math

Read forward, the table budgets your plan; read backward, it is Chapter 8's decoder of authorial intent.

The smell. Any stated constraint — $n \leq 20$, $n \leq 10^5$, and $n \leq 10^{18}$ are three different problems wearing one statement. "At least how many questions / weighings / comparisons?": an information bound begs to be counted. A plan whose cost is unestimated: estimate before executing.

★★ • Meet in the Middle

How many subsets of $\{1, 2, 3, 4\}$ sum to 5 — and how would you count sub-set-sums among 40 numbers, where brute force over $2^{40} \approx 10^{12}$ subsets is dead?

Pause. Split the problem, not the difference.

Solution. Halve: $L = \{1, 2\}$, $R = \{3, 4\}$, with subset sums $\{0, 1, 2, 3\}$ and $\{0, 3, 4, 7\}$. A full subset is a pair, one part from each side, and it hits 5 when $s_l + s_r = 5$: the pairs $(1, 4) = \{1\} + \{4\}$ and $(2, 3) = \{2\} + \{3\}$. **Two subsets**, matching direct enumeration. For $n = 40$: enumerate each half's $2^{20} \approx 10^6$ sums, sort one side, binary-search the complements — about $2^{20} \cdot 20$ operations. Seconds instead of centuries. The only idea is arithmetic: 2^n is lethal, $2 \cdot 2^{(n/2)}$ is cheap, and pairs of half-subsets biject with subsets. Budget arithmetic did not check the method — it *invented* it. ■

The Move — Buy half the exponent. When 2^n kills you, ask whether the problem splits into halves that meet at a seam.

The information-theoretic face already won a battle in Chapter 7: the twelve-ball weighing had 24 possible answers and a three-outcome scale — $3^2 = 9 < 24$, so two weighings are dead; $3^3 = 27$ just clears. And the bound did double duty, *designing* the scheme: each weighing must split the survivors into near-equal thirds, hence four against four, while six against six wastes the balanced outcome. Count the answer space, count the channel capacity, let the ratio steer — the same reflex prices twenty questions ($2^{20} \approx 10^6$ objects), comparison sorting ($\log_2 n! \approx n \log n$), and, read backward, the setter's constraints.

Machines and this move. Asymptotics are the machine's element, yet the pre-computation estimate is a planning behavior automated agents notori-

ously skip — and the information bound is a proof about *all* strategies, not a strategy: that nonconstructive flavor again.

Three Set Pieces

Three famous stories, each a tool in its best-known costume — and each with a correction the retellings usually skip.

★★ • The Seven Bridges of Königsberg

Königsberg, 1736 — now Kaliningrad. The river Pregel divides the city into four land masses — the island of Kneiphof, the north bank, the south bank, the eastern spit — joined by seven bridges. Can a stroll cross every bridge exactly once?

Pause. You have the tool now: name the vertex.

Solution. Euler's move — arguably the founding act of graph theory and of topology, "the geometry of position" in his phrase — was to *delete the city*: each land mass a dot, each bridge a line. Only the incidence structure survives. Then the observation: a walk passing through a vertex consumes its bridges *in pairs*, one in and one out, so every land mass except possibly the start and the end needs an even number of bridges. Königsberg's degrees are 5, 3, 3, 3 — summing to $14 = 2 \times 7$ bridges, the handshake lemma's quiet sanity check — four odd vertices, where a walk tolerates at most two. No such stroll exists. ■ Euler proved the impossibility direction; that at most two odd vertices *suffices* on a connected multigraph was nailed down by Hierholzer in 1873.

The moral is this chapter's opening move in miniature: the insight was not about bridges. It was the decision that only structure matters — one abstraction, then parity a child can check.

The Move — Delete the city. Strip every feature the question does not consume; what remains is usually a solved problem.

★★★ • The Circle of Josephus

The legend: Yodfat, 67 CE — the historian Flavius Josephus, trapped with his soldiers, survives a death circle. Honesty first: his own account (*The Jewish War*, book 3) says lots were drawn, and that he remained at the end "by chance or by providence." The deterministic version — n people in a circle, count around, every k -th falls, the last one lives — is the later mathematicians' formalization. Where should you stand?

Pause. Solve the two-person circle, then the three.

Solution. Let $J(n)$ be the survivor's position, 0-indexed, with step k . After the first elimination (position $k - 1$), the remaining $n - 1$ people play *the same game* shifted by k : **$J(n) = (J(n-1) + k) \bmod n$, with $J(1) = 0$.** One line, $O(n)$ to run. Verified by simulation for $k = 2$: $n = 5$ eliminates 2, 4, 1, 5, leaving **3**, and the recurrence gives 0, 0, 2, 0, 2 $\rightarrow$ position 3 ✓; $n = 6$ eliminates 2, 4, 6, 3, 1, leaving **5**, and the recurrence gives 4 $\rightarrow$ position 5 ✓. For $k = 2$ there is a closed form: write $n = 2^m + L$ with $0 \leq L < 2^m$; the survivor, 1-indexed, is **$2L + 1$** . Equivalently — a jewel — *rotate n 's binary representation left by one bit*: $41 = 101001_2 \rightarrow 010011_2 = 19$, and indeed $41 = 32 + 9$ gives $2 \cdot 9 + 1 = 19$ ✓; $6 = 110_2 \rightarrow 101_2 = 5$ ✓. (The first pass eliminates the evens, and the survivors relabel as exactly a binary shift — recursion made visible in the radix.) The legend's numbers: with $n = 41$ and $k = 3$, the recurrence run in full gives $J(41) = 30$ — position **31**, counting from 1, with the second-to-last survivor at **16**. The traditional telling, Josephus at place 31 and a companion at 16, is arithmetically exact — whether the historian computed or improvised. ■

The Move — Shrink the circle. One elimination turns the game into a smaller copy of itself, relabeled; solve the smallest and lift the answer back.

★ • The Schoolmaster's Sum

The story: Büttner the schoolmaster sets his class to summing $1 + 2 + \dots + 100$, expecting a quiet hour; little Gauss writes 5050 on his slate almost immediately. How — and did it happen?

Pause. You met the move in Chapter 9.

Solution. Pair first with last: $1 + 100, 2 + 99, \dots, 50 + 51$ — fifty pairs, each totaling 101: $50 \times 101 = \mathbf{5050}$. The robust general form dodges any worry about odd n : write the sum forwards and backwards and add columnwise — $2S = n(n + 1)$, so $S = n(n + 1)/2$. ■ Now the record: the anecdote's earliest source, Sartorius von Waltershausen's memorial biography of Gauss (1856), mentions *an* arithmetic-series task but names neither 1-to-100 nor the pairing trick; the vivid details accreted over a century of retellings, and Brian Hayes (*American Scientist*, 2006) traced dozens of mutually contradictory versions. Tell it as legend; teach it as mathematics. The real content is the involution $k \mapsto n + 1 - k$ that makes every pair constant — bedrock that sums any arithmetic series and shows $C(n, k) = C(n, n-k)$.

The Move — Pair to a constant. Chapter 9's symmetry tool in its Sunday clothes.

The library, complete

Sixteen tools across three chapters — boxes invented, charges conserved, extremes interrogated, mirrors held, colorings engineered; hypotheses strengthened, histories compressed, optima morphed, thresholds bisected, tables run backward; vertices named, symmetries quotiented, years turned into n , sums counted sideways, worlds averaged, budgets read. Every volume in this library is a way of *seeing*, not of grinding. The machines have read the library too — they have read everything. Where they now execute it faster than any human, and where they still cannot reach for it unprompted, is the next Part's question. It comes with dates and numbers.

PART IV

The Edge

CHAPTER 12

The Scoreboard

Parts I through III of this book were written in a library. This part comes from a front.

Between December 2024 and July 2026, a run of contests that had stood for years — one of them a benchmark built precisely to outlast the machines — were decided, and not in our favor. Software now outscores people at several of the things the previous eleven chapters taught you to love. There is no honest way to write a book about natural intelligence in 2026 without putting that on the table first, in full, with dates.

So Part IV is a war report in three dispatches. This chapter is the ledger of what fell. Chapter 13 walks the moats that held. Chapter 14 covers the contested ground in between, where the line was still moving as this book went to press.

Three rules govern all three chapters, because this part carries a special obligation: it will be read after some of its numbers have gone stale, and it intends to be respected anyway.

Every claim is dated. "As of July 2026" is not a verbal tic; it is the load-bearing phrase of the whole part. A capability claim without a date is advertising.

Every capability is tagged. **FALLEN** means machines match or beat people on the dated evidence. **ERODING** means the human lead is real but shrinking on a measured clock. **DURABLE** means a five-year human lead is still a defensible bet — defensible, not sacred.

Claims that do real work carry a grade. ♀ marks a primary-sourced result, △ a secondary or contested figure, ⚙ a mechanism — why something happens — and ♦ a rule of thumb. Where a number comes from a leaderboard aggregator rather than a lab, the text says so.

The tags do work that adjectives cannot. "AI is amazing" and "AI is overhyped" are both moods; **FALLEN**, dated December 2024, is a checkable statement about the world. This Part trades moods for statements.

If you are reading this in 2028 and some line below has been overtaken, good. That is what the dates are for. Judge the method, not the snapshot. And no cheering in either direction: a scoreboard that flatters the home team is not a scoreboard.

The test built to outlast machines **FALLEN**

Chapter 3 ended with the Abstraction and Reasoning Corpus — the Sphinx's gate rebuilt for machines. Here is what happened to it.

ARC-AGI-1 (2019) made a clean bet: small grids of colored cells, a few examples of a transformation, then a new case — find the rule. No trivia to memorize, no training set to mine; each task asks for a fresh abstraction. The design goal was never to be hard the way chess is hard; it was to measure how efficiently a solver acquires a brand-new skill from a handful of examples — the thing children do all day and machines, for years, could not do at all. The bet held for five years, through a fifty-thousand-fold scale-up of the language models thrown at it ∞ . For half a decade the machine scores sat near the floor while nearly everything else in AI moved. Then, in December 2024, OpenAI's o3 — spending heavily on computation at test time, adapting to each task on the fly — reached 87.5% in its high-compute mode. Five years of nothing, then everything inside a season.

The successor was built harder. ARC-AGI-2 (March 2025) was calibrated on more than 400 human testers in San Diego: every evaluation task is solvable, within two attempts, by at least two ordinary people, and the average individual test-taker scores about 60%. At launch, pure language models managed 0–4%. Sixteen months later — July 2026 — the frontier leaderboard showed roughly 85% (GPT-5.5), an aggregator figure Δ , above the average human and at the grand-prize bar.

Now the asterisk that matters: cost. On the competition's efficiency track, where entrants get about \$50 of compute for 120 tasks, the best score was 24.03% (NVARC, December 2025) ∞ , and the private-set grand prize re-

mained unclaimed as of July 2026. Machines can now do this kind of abstraction; doing it for pennies is another matter. A person solves these puzzles on a sandwich. That thread — watts, words, the economics of thought — is picked up in Chapter 13, because it is one of the things that held.

The verdict, stated precisely: static abstract reasoning, as a capability, fell by July 2026. The efficiency argument survives. And the benchmark's own designers moved the flag to interactive worlds — ARC-AGI-3 — where the score inverts spectacularly. That battle belongs to Chapter 14.

One year after "humanity has prevailed" FALLEN

Competitive programming fell in two acts, twelve months apart, and the dates deserve to be written out in full.

Act one: July 2025, the AtCoder World Tour Finals, heuristic division — open-ended optimization against the world's best. The human champion, who competes under the handle Psycho, beat OpenAI's entrant and posted three words that traveled around the world: "humanity has prevailed."

Act two: Tokyo, July 9, 2026. At the same finals one year on, an OpenAI model swept all five problems and scored 8,300 points; the best human managed 4,300 Δ (press-reported figures). Not edged out — nearly doubled. The same year's ICPC World Finals, the premier team contest for university programmers, had already gone the same way: an OpenAI system solved all twelve problems (2025).

Note what kind of contest Tokyo was. The heuristic division has no single right answer — entrants chase the best score they can find in a vast search space. It fell anyway, and that matters for the pattern at the end of this chapter: what makes a task machine-winnable is not a unique answer but automatic scoring. An open-ended problem with a scoreboard is still a closed problem.

The lesson is not that the 2025 victory was hollow. It was real, and it was earned against the best system then built. The lesson is the interval. The distance from "humanity has prevailed" to a clean sweep was exactly one

year. Keep that interval in mind whenever this book says **ERODING**: the AtCoder door is what eroding looks like from the inside — a genuine human win, honestly celebrated, with a countdown attached that nobody could read.

The proof line **FALLEN**

Olympiad mathematics — the summit Part III trains toward — held longer than most predictions and then fell on schedule with the rest.

July 2025: two systems, DeepMind's Gemini Deep Think and an experimental OpenAI model, reached the International Mathematical Olympiad's gold-medal threshold of 35 points out of 42 $\approx$. Gold-level performance on the world's most selective mathematics contest, from machines, within the working lifetime of everyone who ever said it was impossible.

It did not stop at contests. By 2026, frontier systems had settled 9 of 353 open problems from Erdős's famous collection, with the proofs checked in the Lean formal-verification system, and had disproved Erdős's planar unit-distance conjecture (announced 2026) $\triangle$ (press-reported). An amateur working with GPT-5.4 Pro closed a problem about prime sets that had stood open for sixty years (2026) $\triangle$.

Read that last sentence again: an amateur. The machine supplied technique; the person supplied the question, the persistence, and the judgment of what counted as finished. Across the 2025–26 mathematics results, the human role narrowed to three things — choosing the problem, setting the strategy, and verifying the result. File those three away. They are the skeleton of the next chapter.

And the Lean detail is not a technicality; it is the mechanism on display $\otimes$. A formal proof checker is the cheapest, strictest verifier there is — a referee that works for free, never tires, and cannot be charmed. Give a searching machine that referee and it can try a million dead ends a night and keep only what compiles. Mathematics did not fall despite being rigorous. It fell because it is rigorous.

As this book went to press, IMO 2026 was underway in Shanghai (July 10–21, 2026), with prediction markets pricing an AI gold at about 82% $\triangle$. The result lands on Chapter 14's watch list; this book states its bets before the results come in, not after.

The white coat, narrowly FALLEN

Clinical diagnosis was supposed to sit safely behind licensure and lived experience. On one specific version of the task, it did not. A Harvard/Beth Israel Deaconess study (April–May 2026) ran an o1-class model against real emergency-department records: 67.1% diagnostic accuracy for the model, versus 55.3% and 50.0% for two groups of expert attending physicians ∇ (press-reported study). The researchers' own epitaph for the exam era: "we're already at the ceiling" of multiple-choice benchmarks.

That sentence deserves a slow read. The multiple-choice exam — the instrument a century of professional licensure is built on — stopped being able to distinguish machine from expert somewhere before mid-2026. Whatever we still want to claim about human judgment in medicine, the exam is no longer the place we can claim it.

Now the width of the claim, exactly. The model read records; it did not run a department. The study was text-only. The same class of systems tends to over-order tests. And the gap between a correct diagnosis on paper and a treated patient in a working hospital remains wide (Stanford–Harvard State of Clinical AI, 2026). What fell is reading a prepared chart better than the median expert reader — a static, fully specified, checkable task. Hold on to those three adjectives. They are about to become this chapter's thesis.

The rest of the ledger FALLEN

Six more entries, compressed into a table because this Part respects your time. Each is real; each carries its own asterisk — and the asterisks are where the next two chapters live.

What fell	Dated evidence	The asterisk
Expert-exam recall and reasoning	Humanity's Last Exam: best ~53.3%, frontier models tightly clustered (July 2026) $\triangle$	the exam is saturating as a differentiator between models — no model has aced it
Personalized persuasion	personalized GPT-4: +81.2% odds of shifting opinions versus human debaters (Nature Human Behaviour, May 2025) ss	persuading is not the same as answering for what you persuaded someone to do — Chapter 13
Theory-of-mind test batteries	matched adult humans up to 6th-order reasoning (Frontiers in Human Neuroscience, 2025) ss	static vignettes; adapting behavior to a mind tracked over time still fails (DialToM, 2026)
Author-voiced prose, expert-judged	fine-tuned models preferred by expert readers 62% of the time (CHI, April 2026) ss	the full story — and what survives it — is Chapter 13's centerpiece
Drug-design execution	AlphaFold3 $\rightarrow$ IsoDDE (February 2026); first AI-designed oncology compounds targeted for human trials in 2026	zero FDA approvals as of mid-2026; designing a molecule is not the same as being allowed to exist
Driving, geo-fenced	Waymo: 220.6 million rider-only miles with 91% fewer serious-injury crashes than matched human drivers (March 2026) ss	a narrow, mapped, data-industrialized domain; stopped school buses and dead traffic lights (2025–26) still break it — the general case is Chapter 13's first moat

The pattern of falling

Lay the fallen side by side and one profile emerges. Contest programs. Olympiad proofs. Abstraction grids. Exam questions. A prepared patient chart. Debate prompts. Prose in a known author's voice. Every one of them is:

Static. The task sits still while you work on it. The chart is already written; the problem statement does not fight back; no supplier lies to you halfway through. Time is not an opponent.

Well-specified. The goal arrives from outside, complete. Prove this. Diagnose this. Maximize this score. Nobody on the fallen list had to decide what was worth wanting — the hardest move in the setter's craft (Chapter 8) had already been made for them.

Verifiable. A machine can check the answer cheaply and exactly. The proof compiles in Lean or it does not. The program scores 8,300 or it does not. The grid matches or it does not.

The third property is the engine ⚙️. Where checking is cheap, training is possible: generate a million attempts, keep what the checker blesses, repeat until the checker is always pleased. A cheap verifier converts raw computation into skill. That is why the theorem — the crown jewel of human reasoning, the summit Part III treats with reverence — fell before the unfamiliar kitchen did: a proof can be machine-checked; breakfast cannot. Chapter 14's vending-machine result makes the same point with money.

There is an inversion of prestige buried here, and it stings. Humans ranked these tasks by how hard they are for us: the olympiad above small talk, the proof above the kitchen. Machines sort tasks along a different axis entirely — how hard they are to specify and to check ⚙️. The two rankings are nearly perpendicular. That is why the fall order looks upside down from inside a university and exactly right from inside a training run. Moravec pointed at this tilt decades ago; 2024–2026 measured it.

Three corollaries, each earned by a dated example:

Falls are sudden. ARC-AGI-1 sat near the floor for five years, then hit 87.5% inside a season (December 2024). AtCoder went from human champion to clean sweep in exactly twelve months (July 2025 → July 9, 2026). Capabilities arrive as step functions more often than ramps, and a long quiet before a fall tells you almost nothing about the date of the fall ⬆️. The suddenness has a mechanism ⚙️: capability accumulates invisibly — in scale, in tooling, in test-time tricks — while the score sits still, like water rising behind a

dam. The score measures the spill, not the reservoir; when the wall is finally topped, the jump looks instantaneous, though the filling never was.

Falls are one-way. Nothing on this list has un-fallen. Benchmarks get retired and rebuilt harder — ARC-2 after ARC-1, ARC-3 after ARC-2 — but no demonstrated capability has been lost. The scoreboard is a ratchet. Any plan that needs a fallen capability to un-fall is not a plan ♦.

A fallen benchmark is not a deployed replacement. The ~85% ARC score (July 2026) ran on a budget far beyond the \$50 that bought 24.03% (December 2025). The 67.1% diagnosis figure came with over-ordered tests and no ward to answer to. The designed drugs had zero approvals as of mid-2026. Capability falls first; economics, integration, and liability fall later, or not at all. A good deal of Chapter 13 lives in that gap.

One more thing belongs in an honest ledger, and it is not a number. If something on the fallen list was *your* summit — the skill a life got organized around — this book will not talk you out of the sting. The best heuristic programmers alive were beaten at the thing they had built their lives around, in public, twelve months after their champion's finest hour. That is real, and it is not the end of the analysis: what we loved about these skills was rarely the checkable part, and the checkable part is all that fell. The next chapter makes that claim precise.

So, a reader's triage — three questions to put to any task, including the one you are paid for ♦:

1. Does the task sit still while you work on it?
2. Is the goal handed to you, fully specified?
3. Can a machine check the answer for pennies?

Three yeses is the profile of everything that fell by mid-2026. Every no is a moat — and a moat, remember, is not a wall. It is ground that stays defensible while you stand on it and work.

The moats are next.

SOURCES

All accessed July 16, 2026. Press and aggregator items are marked; treat their figures with the confidence noted in the text.

- ARC Prize Foundation — ARC-AGI-2 specification and human calibration: arcprize.org/arc-agi/2
- ARC Prize 2025 results and analysis (December 2025): arcprize.org/blog/arc-prize-2025-results-analysis · technical report: arxiv.org/html/2601.10904v1
- ARC-AGI-2 frontier score ~85%, GPT-5.5 (July 2026) — aggregator, secondary: benchlm.ai/benchmarks/arcAgi2
- AtCoder World Tour Finals 2026 sweep, 8,300 vs 4,300 (July 9, 2026) — press: the-decoder.com/openais-ai-beats-every-human-at-atcoder-a-top-competitive-programming-contest/
- International Mathematical Olympiad 2026, Shanghai (July 10–21, 2026): imo-official.org/editions/2026/
- Erdős problems settled and conjecture disproved (announced 2026) — press: scientificamerican.com/article/ai-uncovers-solutions-to-erdos-problems-moving-closer-to-transforming-math/
- Harvard/Beth Israel Deaconess emergency-diagnosis study (April–May 2026) — press: fortune.com/2026/05/04/harvard-study-ai-outdiagnose-doctors-openai-o1-preview/
- Humanity's Last Exam leaderboard (July 2026) — aggregator, secondary: llm-stats.com/benchmarks/humanity-s-last-exam
- Conversational persuasiveness of personalized GPT-4, Nature Human Behaviour (May 2025): nature.com/articles/s41562-025-02194-6
- Higher-order theory of mind, Frontiers in Human Neuroscience (2025): frontiersin.org/journals/human-neuroscience/articles/10.3389/fnhum.2025.1633272/full
- DialToM (2026): arxiv.org/pdf/2604.20443
- Fine-tuned AI writing versus MFA-trained experts, CHI (April 2026): dl.acm.org/doi/10.1145/3772318.3791276
- Isomorphic Labs first-in-human AI-designed drug trials (2026) — press: mlq.ai/news/isomorphic-labs-launches-human-trials-for-ai-designed-drugs/
- Waymo Safety Impact, 220.6M rider-only miles (March 2026): waymo.com/safety/impact/

CHAPTER 13

The Moats

Everything in the last chapter fell in about twenty months. It would be reasonable to open this one braced for consolation prizes — the soft, unfalsifiable comforts about creativity and the human touch that fill airport books on this subject.

This chapter offers none of that, because it was compiled under an anti-romantic rule: every human edge below was attacked before it was admitted. Each candidate was hunted for counter-evidence, dated, and — where the evidence went the other way — demoted to Chapter 12 or 14. What survived the hunt is short, specific, and far better defended than the folk version ever was.

The tag **DURABLE** here means exactly this: on the evidence as of July 2026, a five-year human lead is a defensible bet. Defensible — not guaranteed, not sacred. Five moats qualify. Then comes the flip: a result about expert taste that this chapter refuses to soften, because the moats are only credible if the losses are counted at full price.

#	Moat	The load-bearing evidence, dated
1	The world pushes back DURABLE	video models at chance on impossible physics, humans >95% (IntPhys 2, Meta, 2025)
2	Twenty watts DUR-ABLE	<100M words by age 13 vs 1,000–10,000× more; ~20 W vs 100–750 MW campuses (2026)
3	Learning that sticks DURABLE	continual learning named the consensus #1 missing capability (2026)
4	Knowing you don't know DURABLE	no individuated metacognition (2026); 1,313 court cases of AI fabrication by April 2026
5	The signature DUR-ABLE	liability attaches "regardless of who drafted it" (2026); EU human-oversight duties from August 2026

Moat 1 — The world pushes back DURABLE

Start with the cleanest measurement. Meta's IntPhys 2 benchmark (2025) shows video models judging physically impossible events — the kind that violate object permanence, solidity, continuity — at chance level, roughly 50–60%. Humans score above 95% $\approx$. The physics a one-year-old enforces before acquiring a single word is, as of 2025–26, missing from systems that write passable sonnets. This is Moravec's paradox — hard things easy, easy things hard — and it remains the best-supported human edge on the books.

The hardware side tells the same story with money attached. Robot dexterity is genuinely moving: 1X's NEO humanoid carries 22-degree-of-freedom tendon-driven hands; Physical Intelligence's $\pi 0.7$ foundation model (April 2026) folds laundry and assembles boxes; its $\pi^*0.6$ line learns from real-world reinforcement. And yet the same 2026 technical literature calls the field "nowhere near human capabilities" (Construction Physics, 2026; a running Science Robotics debate, 2026). Deployment numbers agree: Figure's 02 robot logged about 1,250 operating hours across eleven months at a BMW plant; Tesla missed its 2025 Optimus production target by roughly 90%, with Musk conceding on the Q4 2025 earnings call that the robots were "not doing useful work" yet; credible estimates put productive autonomous humanoids at hundreds to low thousands worldwide by end-2026. Not zero. Also not a workforce.

Why this moat holds, mechanically $\otimes$: two structural reasons, not one hopeful one. Hardware iterates on physical time — a gripper that breaks cannot be patched tonight and pushed to a million users tomorrow. And physical training data is scarce: there is no internet of touch. The field calls this the physical data wall. Text fell first because centuries of text were already piled up, labeled, and cheap to check; the world hands over its data one grasp at a time.

Now the honest exception, because this moat has a hole with a company logo on it. Waymo's driverless fleet passed 220.6 million rider-only miles by March 2026, with 91% fewer serious-injury crashes than matched human drivers $\approx$ — in a domain narrowed by geo-fencing, high-definition maps, and a decade of industrialized data collection. Even there the long tail bites:

stopped school buses and dead traffic lights produced failures through 2025–26. The rule the exception teaches ♦: where a physical domain can be fenced, mapped, and data-farmed, it falls like a benchmark. What stands is the open world — the unfamiliar kitchen, the cluttered job site, the task nobody fenced in advance. General-purpose embodied common sense is a defensible five-year bet as of July 2026. A five-year bet on any particular fenced-off slice of it is not.

Moat 2 — Twenty watts and a hundred million words

DURABLE

Two numbers, and neither has moved despite years of well-funded trying.

First: a child hears fewer than 100 million words by age thirteen. Frontier language models train on three to four orders of magnitude more — a thousand to ten thousand times the linguistic experience — and still need it (BabyLM 2026 call for papers). The BabyLM challenge has spent four years hunting for training tricks that close the gap at child scale; the results help at small scale and approach nothing like a child's sample efficiency on open-ended learning ∞.

Second: the brain runs on about 20 watts. Per watt, published estimates put biological cognition somewhere between a thousand and a million times more efficient than current AI hardware. Meanwhile, single AI campuses draw 100–750 megawatts (2026) — the top figure is the resting draw of tens of millions of human brains — and more than 10 gigawatts of new AI capacity was announced in 2025 alone.

Read the two numbers together and the point sharpens ⚙: this gap is architectural, not a matter of scale, because brute scale is precisely what the gap measures. Every added gigawatt widens it. Whatever brains do to pull deep structure out of thin data has not been replicated, and it is not for lack of budget.

One honest qualification. In-context learning — a few examples shown inside the prompt — narrows the practical gap for the task at hand, and it genuinely works. But it evaporates when the session ends. Few-shot per-

manent learning did not exist as of July 2026. That distinction is a moat in its own right, and it gets the next section.

Why this belongs in a puzzle book: sample-efficient learning is the puzzle habit. One rope, one fuse, one properly digested solution becomes a Move you own for the rest of your life. That efficiency — a single example turning into a permanent skill — is exactly what Parts II and III train, one puzzle at a time.

Moat 3 — Learning that sticks DURABLE

Ask the field itself, in 2026, which single capability is most conspicuously missing, and the answer is unusually consistent: continual learning. It headlines "AGI's Last Bottlenecks" (AI Frontiers, 2026); it is why Karpathy called this a "decade of agents" rather than a year (2025); it fuels the "continual-learning overhang" debate (2026) ♾. The field's own clinical image: models wake up with anterograde amnesia.

Here is what that means in practice. Everything a model learns inside a session — your corrections, its own discoveries, the bug it finally understood — evaporates when the session ends. The colleague who was brilliant yesterday arrives today having never met you. Every morning, the first morning.

The workarounds are real, and they are not learning ⚙. Persistent agents keep external memory — wiki files, notes-to-self, retrieval stores — and these help the way a stranger's notebook helps the stranger who inherits it: the skill never moves into the system itself; it sits in a file that the next session may or may not read correctly. Meanwhile the mirror-image problem gained production teeth: catastrophic forgetting — train in the new, damage the old — became a working engineer's headache in 2025–26 as agents became persistent services ♾.

Against all this, the human baseline barely needs stating: skills compound across decades. Your ten-thousandth conversation informs your ten-thousand-and-first. A reader who absorbs this book's Moves in 2026 will still own them in 2046, no retrieval store required. Chapter 15 builds directly on this asymmetry — the mistake journal, upsolving, spaced review — because a compounding learner should invest where compounding pays.

Grade it honestly, though. This is the single most-attacked research target in AI $\triangle$: robots already learn from real-world experience in narrow settings (π 0.6, 2026), *and the incentive to crack the general case is measured in trillions*. The **DURABLE** here is pinned to today's architectures*, and today has a date on it: July 2026. If one moat in this chapter falls to an architectural change rather than to scale, expect it to be this one.

Moat 4 — Knowing you don't know **DURABLE**

Start with the part that is eroding, so the durable part earns your belief.

Raw error rates are collapsing. Hallucination in grounded summarization: 3–8% in 2023, down to 1.0–2.5% by 2026 (HHEM figures). Faithfulness errors in retrieval-augmented setups: 4–9%. Closed-book factual error on Simple-QA-style tests: roughly 4–10% for frontier models, with extended-thinking modes roughly halving errors (April 2026; aggregator figures) $\triangle$. OpenAI's September 2025 paper "Why Language Models Hallucinate" argues the root cause is benchmark incentives that reward guessing over abstaining — an argument that the whole phenomenon is fixable $\triangle$. Anyone building a human moat out of AI error rates is building on sand, and this book declines to.

The moat sits one layer down, and it has not moved. Studies through 2026 find language models show no signs of individuated metacognition ∇ : the model's stated confidence does not track what that particular model actually knows, and its willingness to say "I don't know" turns out to be a prompt artifact — rephrase the "Unknown" option and the abstention rate shifts, a bias the model cannot detect in itself (arXiv, 2026). A model can be prompted to *perform* doubt. It does not *have* doubt ∇ . The difference sounds philosophical, right up until it meets a courtroom.

So meet the courtroom. By April 2026 a running tracker had counted 1,313 documented court proceedings involving AI-fabricated content — invented cases, phantom citations — 496 of them involving licensed attorneys ∇ . US courts issued more than \$145,000 in sanctions in the first quarter of 2026 alone. Nebraska delivered the first indefinite suspension of a law license over AI-fabricated filings. Note the anatomy of these failures: nothing exot-

ic. A professional asked a fluent system a question, received a fluent answer, and signed it.

The economics land exactly where the case law does ☸: when generation becomes nearly free, verification becomes the scarce input. Calibrated doubt — knowing which of your own answers deserve checking, and saying "I don't know" at the right moments for the right reasons — was, as of July 2026, a human craft. It is also a trainable one; Part V drills it.

Mark the shape of this moat carefully. Its metacognitive core is durable while its outer wall — raw accuracy — erodes every quarter. Do not paraphrase this section as "AI makes things up." It makes things up less every year. It still does not know when it is doing it.

Moat 5 — The signature DURABLE

The Nebraska suspension points at something larger than error rates: someone had to be the one who signed.

The doctrine that crystallized in the 2026 sanctions wave is blunt: a signature "creates liability for its contents regardless of who drafted it" ⚖. The courts did not ask whether the machine was clever. They asked who signed. From August 2026, the EU AI Act's human-oversight duties for high-risk systems write the same assignment into statute. Accountability, it turns out, is not a cognitive capability that machines lack and might one day acquire. It is a role institutions assign to entities that can be sanctioned, disbarred, sued, and shamed ☸. There is no benchmark to climb, because it is not a performance — it is a position. Durable by construction.

Goal-ownership is the same moat seen from the front. Reviews of agentic "AI scientists" in 2026 find deployment running technology-push: the systems execute impressively while problem selection, high-level strategy, and standards of proof remain human-set ⚖ (arXiv reviews and case studies, 2026). Chapter 12's mathematics pattern repeats everywhere: machines supply technique; humans decide what is worth proving and what counts as proven.

Log the erosion at full price, though. Judges trained on 700,000 citation-matched paper pairs can learn something that behaves like scientific taste (March 2026), and one frontier lab reports that in live research sessions its model's proposed next step now beats the human researcher's proposal about half the time (2026; secondary report, ~75% confidence) △. Proposing the next step is not choosing the mountain. But the distance between those two is smaller than it sounds, and this paragraph may age quickly.

Two things stay put even so. First, longitudinal social reasoning: models pass static mental-state tests (Chapter 12) yet fail to adapt their *behavior* to a mind tracked across time (DialToM, 2026) ▴. Second, and deeper: wanting. A system can optimize a goal handed to it; as of July 2026, nothing is at stake for the system itself. It cannot be fired, bankrupted, embarrassed, or proud. Skin in the game is not a metaphor here — it is the mechanism the entire accountability structure runs on ✿. The right to be the one who decides, and the obligation to answer for the decision, is a human monopoly not because machines are dim, but because institutions are made of consequences.

The flip — expert taste and eighty-one dollars

This chapter owes you one story told at full strength, because it is the one that would be easiest to soften.

The setup (CHI, published April 2026): twenty-eight MFA-trained writers — people whose profession is telling good prose from almost-good — judged passages by language models emulating fifty acclaimed authors. Under ordinary prompting, the experts preferred the human-written prose 82.7% of the time. The moat held. Connoisseurship could tell.

Then the researchers spent about \$81 fine-tuning a model on an author's actual corpus. The experts' preference flipped: 62% for the machine. Lay readers had preferred the machine all along, in both conditions. The study's debriefs record expert "identity crisis." The human labor being emulated was estimated at roughly \$25,000 of writing time. The flip cost eighty-one dollars ▴.

Say the unsettling part plainly, because pretending otherwise would poison the rest of the chapter. Expert taste, pointed at typical-quality work, was supposed to be deep — the accumulated, embodied judgment of a career. As a *detector*, it was defeated for the price of a good dinner, and the experts did not merely fail to spot the machine; they preferred it. If a professional edge amounts to "I can tell good from almost-good," that edge is not a moat. It is a fine-tuning run that has not happened yet.

Nor is this confined to prose. OpenAI's GDPval (September 2025, updated since) rated frontier-model deliverables equal to or better than those of professionals averaging fourteen years' experience on about half of 220 real occupational tasks — roughly a hundred times faster and cheaper.

Now what survives — because something does, and it is better evidenced than the thing that fell.

The tails. Pools of human writing are 2–8× more semantically diverse than pools of model writing, and the homogenization persists even after deliberate diversity-boosting interventions (Doshi & Hauser, *Science Advances*, 2024; further results 2025–26) ♢. Models regress to the cultural center ♠ — they are, nearly by construction, engines of the expected. The center is exactly where they now beat the median professional. But the tails are where new genres, new markets, and new science come from, and the tails remained human as of 2026. One excellent machine writer is impressive; a million of them are, collectively, the same writer.

The choosing. Taste-as-discrimination fell at the center. Taste-as-selection — which problem deserves your year, which standard of proof counts, which of a thousand adequate drafts is worth staking a name on — runs through Moat 5, and the 2026 evidence still finds it human-set ♢.

One caveat and one instruction. The caveat: outlier taste is durable partly because nobody has built a benchmark for it yet, and protection by unmeasurability is a weaker kind of safety than protection by evidence △ — expect this section to be stress-tested before most others. The instruction, for anyone whose work is judgment: move your value off the center ♦. Median discrimination is falling. Oddness, selection, and the willingness to sign are not — as of July 2026.

What the moats share

Stand back from the five moats and the flip, and Chapter 12's fallen list reads like a photographic negative of this one.

What fell was static, fully specified, and cheap to verify. What stands is the opposite in every coordinate: a world that pushes back mid-task (Moat 1); learning that compounds instead of evaporating (Moats 2 and 3); doubt that is owned rather than performed (Moat 4); goals that are chosen and signed for (Moat 5); tails instead of centers (the flip's survivors). The machines mastered the closed loop . Every moat is an open one — no fence around the world, no goal handed down, no free verifier, and consequences that land on the agent who acts.

None of this is a promise. Chapter 12 showed that falls are sudden and one-way; Moat 3 is under direct assault by the best-funded research effort in history; the flip happened to professionals at least as confident as any reader of this page. What the moats offer is not safety but direction: they mark where human effort still compounds, and Part V spends that information.

Between the fallen and the durable runs a strip of ground that is neither — the front line itself, where the lead is real and the clock on it is public. That is the next chapter.

SOURCES

All accessed July 16, 2026. Press and aggregator items are marked; treat their figures with the confidence noted in the text.

- IntPhys 2, Meta (2025): arxiv.org/abs/2506.09849
- Robot dexterity assessment (2026) — press: construction-physics.com/p/robot-dexterity-still-seems-hard
- Humanoid deployment reality check, Q2 2026 — press: 1023jack.com/news/the-humanoid-robotics-reality-check-q2-2026-pilot-to-production-status/
- Waymo Safety Impact, 220.6M rider-only miles (March 2026): waymo.com/safety/impact/
- BabyLM 2026 call for papers (child data-efficiency gap): arxiv.org/abs/2602.20092
- AI Frontiers, "AGI's Last Bottlenecks" (2026): ai-frontiers.org/articles/agis-last-bottlenecks

- Karpathy on agent memory and the "decade of agents" (2025–26) — press: inovabeing.com/blog/karpathy-llm-wiki-ai-agent-memory-2026
- No individuated metacognition (2026): arxiv.org/pdf/2605.24299 · abstention as prompt artifact (2026): arxiv.org/pdf/2507.16199 and [arXiv 2604.17073](https://arxiv.org/abs/2604.17073)
- OpenAI, "Why Language Models Hallucinate" (September 2025) — cited by title; hallucination-rate figures from HHEM (2026) and aggregator data (April 2026), secondary
- Legal tracker of AI-fabricated content in court — 1,313 proceedings by April 2026: gc.ai/blog/ai-hallucination-legal-cases
- Fine-tuned AI writing versus MFA-trained experts, CHI (April 2026): dl.acm.org/doi/10.1145/3772318.3791276
- GDPval, OpenAI (September 2025): openai.com/index/gdpval/
- "AI Can Learn Scientific Taste" (March 2026): arxiv.org/abs/2603.14473
- Agentic AI scientists — deployment critique (2026): arxiv.org/pdf/2605.08956
- Homogenization of creative diversity: Doshi & Hauser, Science Advances (2024); replication and extension (2025): sciencedirect.com/science/article/pii/S294988212500091X
- DialToM (2026): arxiv.org/pdf/2604.20443

CHAPTER 14

The Race

Chapter 12 counted the fallen. Chapter 13 walked the moats. This chapter reports from the strip between them: capabilities where the human lead is real, measured, and shrinking on a clock anyone can read. Everything here is tagged **ERODING** — which makes this the chapter most likely to be wrong by the time you read it. It says so up front, and it is built to be updated rather than believed: dates on every number, ranges instead of points, and a watch list at the end naming exactly which figures to check against the world.

The deeper purpose is the method. Numbers about AI arrive gift-wrapped in someone's excitement or someone's dread. This chapter's real lesson is how to unwrap them.

The clock **ERODING**

The most useful single ruler for the race is METR's time-horizon metric, and it is worth understanding before quoting ⚙️: take a large set of tasks whose difficulty is calibrated in human working time, and find the task length at which a model succeeds 50% of the time. That length is the model's horizon. It converts "how smart is it?" — unanswerable — into "how long a job can you hand it?" — measurable.

The readings, from the primary source. METR's original study (March 2025) found horizons doubling roughly every seven months. Its Time Horizon 1.1 update (January 29, 2026; 228 tasks) put Claude Opus 4.5's 50%-success horizon at 320 minutes — about 5.3 hours — with a confidence interval of 170 to 729 minutes, and GPT-5's at 214 minutes ⚡. The refit doubling time: 131 days measured from 2023, or 89 days measured from 2024.

Notice the honesty inside the primary source itself: the headline 5.3 hours arrives wrapped in a confidence interval running from under three hours to over twelve. That width is not sloppiness; it is what measuring a moving

frontier with 228 tasks actually looks like. Distrust any secondhand version of this number that quotes the point and drops the interval ♦.

Sit with that pair of numbers a moment. The doubling time itself shrank between the two reports — the curve ran ahead of its own extrapolation within ten months. Whatever else is true, "the trend will simply slow down" has so far been the losing side of this particular bet.

One frontier reading, flagged at its true weight: secondary reports (a February 2026 data fit) place a March 2026 preview model near a 16-hour horizon — but that sits at the edge of what the task suite can measure at all, and METR itself warns that estimates above 16 hours are unreliable, since only 5 of the 228 tasks run that long △ (~70% confidence). When a reading sits at the top of the ruler, suspect the ruler first.

Do the math, then doubt it

Here is the extrapolation everyone performs silently. Do it out loud instead, with both endpoints of the measured range. Anchor: 5.3 hours on January 29, 2026. Formula: horizon = 5.3 h × 2^(days elapsed ÷ doubling time). Working week = 40 hours; working month ≈ 170; working year ≈ 2,000.

Date		At 131-day doubling		At 89-day doubling
January	2026	5.3 h		5.3 h
(measured)				
January	2028 (+2 years)	≈250 h — six working weeks		≈1,600 h — nine working months
January	2031 (+5 years)	≈80,000 h — a forty-year career		≈8,000,000 h — four thousand working years

Read the middle row first. Six working weeks is a serious consulting engagement; nine working months is a funded research project. Even the conservative endpoint, two years out, describes handing a machine the kind of work that structures a professional's whole quarter — at coin-flip reliability, on benchmark-clean tasks, if the trend holds. All three of those qualifiers are doing real work; keep them attached.

Read the bottom row again. Taken literally, it says a model in 2031 completes — at coin-flip reliability — a task that would occupy a human for several millennia. The two-year row is a forecast. The five-year row is a reminder that somewhere between the rows, the assumptions stop describing the world ♦. Exponentials are honest about the near term and hallucinate about the far term; that is not a flaw in this particular curve but a property of the method.

Four corrections before you quote even the near row:

1. **Reliability.** These are 50%-success horizons — coin-flip odds. Horizons at 80% reliability run about four times shorter: 3–4 hours for the best models as of January 2026 ☸. For "dependable," subtract roughly two doublings from everything in the table.
2. **The ruler's ceiling.** The suite thins out above 16 hours (5 tasks of 228). Beyond that, the instrument — not necessarily the phenomenon — runs out ☼. Watch for METR's successor suite before trusting any figure past that line.
3. **Benchmark tasks are cleaner than jobs.** The suite's tasks are self-contained and machine-scored — Chapter 12's fallen profile, by design. Real months contain ambiguity, interruptions, and suppliers who lie. The next section prices that difference in dollars.
4. **Exponentials bend — in both directions.** Every physical growth curve saturates eventually; this one, so far, has bent the wrong way for comfort (seven months to 89–131 days between March 2025 and January 2026). Anyone certain the curve must break, and anyone certain it must hold, is selling something ♦.

Now the arithmetic on the question people actually ask. From 5.3 hours to a 170-hour working month is five doublings, almost exactly. At 89 days per doubling, that lands around April 2027; at 131 days, around November 2027. Demand 80% reliability instead of 50% — two more doublings — and the window slides to late 2027 through mid-2028. Stack on the measurement caveats and the honest sentence reads: a dependable machine work-month is plausible somewhere in 2028–2029, *if the trend holds* — and "if the trend

holds" is the entire bet. Both halves of that sentence were written in July 2026. Check the watch list before repeating either.

A year with consequences **ERODING**

Against the optimism of the horizon curve, weigh the benchmark that made agents run a business.

Vending-Bench 2 (Andon Labs; updated July 2026) hands a model a simulated vending operation for a full business year: stock it, price it, negotiate with suppliers, handle the unexpected, stay solvent. No trick puzzles — just a simulated year of consequences. The best model, Claude Opus 4.7, finished the year with \$10,937 in net worth. A merely "good" human strategy is estimated at about \$63,000 $\approx$. Roughly a six-fold gap — not on a logic exam, but on keeping a tiny business alive.

Where does the money go? Into precisely the gaps the horizon metric is too clean to see $\otimes$: reliability over long stretches, recovering from errors instead of compounding them, and not getting scammed by suppliers. Credit where due: the top models win points partly through persistent price negotiation — a socially fluent skill, and a small echo of Chapter 12's persuasion results. But fluency in the moment is not solvency over the year.

It is worth pausing on why a year of consequences resists what the theorem could not $\otimes$. Chapter 12's engine was the cheap verifier: try a million times, keep what the checker blesses. A business year has no checker mid-course — feedback arrives slowly, tangled with luck, and a wrong call in March surfaces as a hole in the accounts in September. Where feedback is slow, sparse, and consequential, brute search stops compounding, and the patient, error-correcting animal still earns its keep.

Two lines to hold together, both dated July 2026: models improve on this benchmark by about \$799 per month, on a fit of $R^2 = 0.96$ — an unnervingly straight line — and that line crosses the \$63,000 human mark around 2031 if nothing bends. So the same report card carries both messages at once: humans dominate the year-of-consequences today, and the gap is closing at a posted, public rate. That pairing — clear lead, visible clock — is what **ERODING** means in this book.

The gap that embarrasses the machines ERODING

The largest measured human-machine gap in mainstream benchmarking, as of July 2026, is not close.

ARC-AGI-3 (launched March 25, 2026) abandoned static grids for interactive environments with no instructions, no stated goals, no described rules. The agent must explore, build a model of the world from scratch, work out what the goal even is, and then plan toward it. It is a laboratory version of being dropped into an unfamiliar situation — the thing every puzzle in Part II rehearses in miniature.

At launch: humans solved 100% of the environments; every frontier model scored below 1% ♪. Ordinary people, near-perfect; the best machines on earth, a rounding error above zero. By July 2026 the best public snapshot had reached about 8% (GPT-5.6 Sol; aggregator figure, ~80% confidence) △.

Resist the urge to build a shrine here. This book's own sourcing gives the reason: ARC-AGI-1 stood for five years before falling inside a season (December 2024); ARC-AGI-2 stood for about sixteen months (July 2026). The family record argues the third one erodes too — and less-than-1% to roughly 8% in under four months is not a wall standing firm; it is a wall being climbed. Betting that this gap holds five years means betting against the exact trend that killed its two predecessors. Hence the tag: ERODING, with the largest current gap on the books. The ARC Prize 2026 season closes in November 2026; the watch list below states in advance what score would force a rewrite.

Insight, when the world talks back ERODING

The same fault line runs through lateral thinking, and it is worth mapping precisely, because folk wisdom puts far too much of it on the human side.

On static riddle sets, the machines already do well: 76–83% across prompting styles on a 139-riddle benchmark (2026 study). The one-shot "aha" — the lightbulb item on a page — is no longer a reliable human lead, and this book will not pretend otherwise. Where humans still clearly win is *interactive* insight: situation puzzles (the parlor genre where a bizarre scenario must be

cracked by asking the right yes/no questions) studied through LatEval and the ~975-puzzle SPLAT set. There, models show "almost negligible" lateral moves in interaction, defaulting to vertical, deductive grinding when the move that scores is to question the frame itself ♪.

Note that this is the same failure ARC-AGI-3 measures, in different clothes ♪: restructuring a representation *while exploring it* — updating not just your answer but your picture of what the question is, mid-conversation with a world that answers back. Part II called this breaking the frame; Chapter 4 trained it. As of July 2026, the interactive version remains human ground, sitting inside the blast radius of the fastest-moving benchmark in the field. **ERODING** — enjoy it; train it; do not lean your identity on it.

The training implication is direct, and Part V takes it up: practice insight in its interactive form. Not the riddle on the page — the machines have that — but the version where you must buy information with questions, notice that your picture of the problem is wrong, and rebuild it while the clock runs. As of July 2026 that skill is measurably scarce in machines, and it is the most puzzle-like thing left on the board.

How to read a trend without hype

This section is the chapter's actual payload. The numbers above will age; the method for handling such numbers should not. Six rules, each one already exercised on this chapter's own data ♠:

1. **Date every number.** An undated capability claim is advertising. Dated, it becomes a data point you can plot and check. Every figure in this Part carries its month for exactly this reason.
2. **Find the definition before the digits.** Fifty percent of *what*? Preferred by *whom*? The ARC-AGI-2 "~85%" (July 2026) and "24.03%" (December 2025) describe the same benchmark under different compute budgets — a 3.5× disagreement produced entirely by the definition, not the capability.
3. **Extrapolate with both endpoints, never the midpoint.** The measured doubling range, 89–131 days, diverges into a six-fold disagreement within two years (≈250 vs ≈1,600 hours by January 2028). An honest range

that wide is not a weakness of the analysis; it is the analysis. A single confident number would be the lie.

- 4. **Ask what the ruler cannot see.** METR's suite thins above 16 hours; lead-erboards saturate; experts stop being able to tell (Chapter 13). A flat line at the top of a scale is a broken instrument as often as a plateau ⚙️.
- 5. **Separate capability from consequence.** A benchmark falling does not mean a job disappearing: the 67.1% diagnostic score arrived with over-ordered tests and no ward to answer to (Chapter 12); designed drugs had zero approvals as of mid-2026 (Chapter 12); the vending gap stood at roughly \$52,000 (July 2026). Deployment runs on economics, workflow, and liability — Chapter 13's terrain — and lags capability by years.
- 6. **Precommit to what would change your mind.** Decide *today* which num-ber, at which threshold, flips your verdict — then check it on schedule. A forecast without a falsifier is a mood. The list below is this book's own precommitment, in print, where it cannot be quietly revised.

The watch list

Five numbers, checkable by anyone, that will grade this Part's homework. Status as of July 16, 2026:

#	Watch	Where it stood at press time	What would move the verdict
1	IMO 2026, Shanghai, July 10–21 — live as this book went to press	prediction markets pricing an AI gold at ~82%; results due in the last week of July 2026	gold: Chapter 12's "proof line" verdict confirmed on sched-ule · no gold: evidence of year-to-year variance the markets missed
2	ARC-AGI-3 leaderboard through the ARC Prize 2026 close (November 2026)	<1% at launch (March 2026) → ~8% (July 2026)	agents crossing 30–50%: downgrade "largest gap," shorten the interactive-insight half-life
3	METR's next time-horizon release (suc-		confirmed >16 h horizons: the two-year column above is on

#	Watch		Where it stood at press time	What would move the verdict	
			cessor to TH 1.1, saturation addressed)	5.3 h measured (January 2026); ~16 h preview unconfirmed △	track; slower: first real bend in the curve
4	Vending-Bench versus the "good human" line	2 ~\$63K	\$10,937 best; +\$799/month, R ² = 0.96 (July 2026)	a step-change (not slope): consequence-bearing agency eroding faster than linear; crossover ≈2031 if the line merely holds	
5	Robot learning-from-experience (π*0.6 successors)		real-world RL demonstrated in narrow settings (2026)	robots that keep skills across days would attack two Chapter 13 moats at once — embodiment and continual learning	

A reader in 2028 holding this page has everything needed to grade it: five claims, five thresholds, five dates. Where the book turns out wrong, the method that wrote it shows how to correct it — which is the only kind of authority a chapter like this can honestly claim.

Where Part IV lands

Add up the three dispatches.

What fell was the closed loop: tasks that sit still, goals handed over complete, answers a machine can check for pennies — the proof, the contest, the exam, the prepared chart (Chapter 12). What held were the open loops: a world that pushes back, learning that compounds across decades, doubt that is owned rather than performed, goals that are chosen and signed for, the tails rather than the center (Chapter 13). And between them runs this chapter's front line — hours doubling every 89–131 days, a business year still six-fold human, an interactive-reasoning gap of 100% against 8% — every figure dated, every figure moving (as of July 2026).

Now the moral, and it is the reason this Part sits inside a puzzle book rather than a policy report.

Look at the durable column one more time. Learning deeply from a handful of examples. Carrying a lesson forward so it compounds. Knowing which of your own answers to doubt. Restructuring the problem while you are inside it. Choosing what deserves the work, and signing for the result. That is not a list of consolations — it is a curriculum, and it is *this book's* curriculum. One puzzle digested into a permanent Move is Moat 2 in action. The mistake journal and upsolving are Moat 3, practiced. Calibration drills are Moat 4. The setter's mind — asking why this problem, why these constraints — is Moat 5. And every frame this book has broken since the nine dots is the muscle ARC-AGI-3 measures.

None of this was designed as a response to the machines — the canon is millennia older than the transistor. But the machines have, without meaning to, run the audit: of everything the gymnasium teaches, they took the checkable and left the compounding.

The gymnasium is four thousand years old. The machines did not close it; they surveyed it, and marked which halls still matter. The scoreboard will keep moving — the dates are on every number so you can move with it. Part V turns the map into a training plan.

SOURCES

All accessed July 16, 2026. Press and aggregator items are marked; treat their figures with the confidence noted in the text.

- METR, Time Horizon 1.1 (January 29, 2026): metr.org/blog/2026-1-29-time-horizon-1-1/
- METR, Measuring AI Ability to Complete Long Tasks (March 19, 2025): metr.org/blog/2025-03-19-measuring-ai-ability-to-complete-long-tasks/
- Doubling-time refit and ~16 h preview figure (February 2026 data fit) — secondary: x.com/aakashgupta/status/2053207639107174452
- Vending-Bench 2, Andon Labs (updated July 2026): andonlabs.com/evals/vending-bench-2
- ARC-AGI-3, ARC Prize Foundation (launched March 25, 2026): arcprize.org/arc-agi/3 · technical report: arxiv.org/abs/2603.24621
- ARC-AGI-3 frontier snapshot ~8% (July 2026) — aggregator figure, ~80% confidence; no primary publication as of press time
- ARC Prize 2025 results and analysis (efficiency-track figures, December 2025): arcprize.org/blog/arc-prize-2025-results-analysis · arxiv.org/html/2601.10904v1

- SPLAT situation-puzzle benchmark (~975 puzzles): arxiv.org/html/2410.06733v1 · LatEval — cited by name
- Riddle-benchmark results, 76–83% across prompting styles, 139 riddles (2026 study) — cited by description; no public URL recorded at press time
- International Mathematical Olympiad 2026, Shanghai (July 10–21, 2026): imo-official.org/editions/2026/
- Chapter cross-references: diagnosis and drug-approval status (Chapter 12 sources); expert-detection flip and accountability doctrine (Chapter 13 sources)

PART V

The Gymnasium

CHAPTER 15

Training the Edge

Part IV ended with a claim that should sound suspicious: the human **P**edges that hold — calibration, continual learning, efficient acquisition, taste, ownership — are exactly the ones a person can train. Suspicious, because "you can train it" is where thinking books traditionally start lying. Training advice is the one corner of this genre where wishful thinking turns a profit.

So this chapter carries the heaviest density of claim grades in the book. What follows is not a philosophy of self-improvement. It is a short list of practices with evidence behind them, a pointer to the famous practices with autopsies behind them, and a protocol for using a machine as a sparring partner without letting it become a crutch. Chapter 16 turns all of it into a schedule.

The gym that does not work

Begin with the graveyard's freshest stone.

The pitch is familiar: ten minutes of app puzzles a day, and memory, attention, and general intelligence rise like a trained muscle. The pitch was tested properly, once, at scale. In 2010 Adrian Owen and colleagues ran 11,430 adults through six weeks of online brain training and measured what moved (*Nature*, 2010). Scores on the trained tasks climbed steeply. Scores on everything else — including untrained tasks that were close cousins of the trained ones — did not move $\emptyset$. A 2016 review by Daniel Simons and six colleagues read the industry's entire evidence file and reached the same verdict: strong evidence of improvement on the practiced games, little for near transfer, none convincing for "your life improves" $\emptyset$. The same year, the US Federal Trade Commission fined Lumosity's maker two million dollars for deceptive advertising. The full autopsy — with five sibling myths in adjoining plots — is in Appendix B $\triangle$.

Why it fails is a mechanism worth owning ☸: practice is brutally specific. You improve at what you literally do, plus a thin ring of neighbors. The classic demonstration is fifty years old. Chess masters can glance at a mid-game position and reproduce it almost perfectly; club players cannot. Scramble the same pieces into random nonsense and the masters' superpower evaporates (Chase and Simon, 1973 ☸). The masters never had better memory. They had a library of chunks, and random boards contain no chunks. The skill was never general; it only looked general from outside.

And here is the corollary this book cannot dodge: puzzles obey the same physics. Solve the fifty in this canon carelessly and the main thing you will get better at is those fifty puzzles. A puzzle book that promises more, with no mechanism, is a brain-training app with better typography.

This book was built around the escape clause.

The escape clause: name the move

The escape clause has an experimental pedigree. In 1980 Mary Gick and Keith Holyoak gave people Duncker's radiation problem — destroy a tumor with rays that wreck healthy tissue at full strength — after having them read a military story with the identical skeleton: a general splits his army into small columns that converge on a fortress from all sides. About one solver in ten cracks the tumor unaided. Having just read the fortress story, still only about three in ten make the leap on their own — the analogy sits in memory, inert. Told "that story you read may help," most solve it at once ☸. The knowledge was there. The recognition wasn't. In the follow-up work, one group transferred spontaneously at much higher rates: people who had read *two* analogous stories and written down the skeleton the stories shared ☸.

That is this book's whole theory, in one experiment ☸. Transfer follows the named, compared schema — not the exposure. Solving the two-ropes puzzle teaches you the two-ropes puzzle. Solving it and then writing "the move was *exploit rates, not lengths* — find the quantity the mess can't touch" files something reusable. A move you can name is a move you can notice, and noticing is the entire game. Part III made the argument in full: expertise

is a library of recognition-action pairs, compiled from solved and *failed* problems, converting search into lookup. A medalist carries perhaps two hundred. This book issues thirty-two — Appendix A — and the difference between reading them and owning them is the training below.

What transfers, then, is move-recognition across varied surfaces ♪ — and it transfers under known, unglamorous conditions that half a century of memory research keeps re-confirming:

Principle	What it means here	Grade
Retrieval beats re-reading	Re-solve from the bare statement; don't reread solutions and nod along	♪ Roediger & Karpicke, 2006
Spacing beats massing	Return after days and weeks, not minutes	♪ Cepeda et al., 2006
Variation beats repetition	Same move, different costume, on purpose	♪ interleaving literature
Difficulty is a feature	Effortful, error-prone practice outlearns smooth practice	♪ Bjork's "desirable difficulties"
Fluency is a liar	"That felt easy" measures familiarity, not learning	♪

The last row is the dangerous one. The practice that *feels* productive — re-reading elegant solutions, watching someone else be clever — produces fluency, and fluency feels exactly like skill from the inside. The practice that works feels worse while it works. Plan for that, or the plan fails.

The instruments

Competitive mathematics and competitive programming form the most refined training culture our species has built for reasoning — unusually honest because its feedback is unfakeable: the proof works or it doesn't; the code passes or it doesn't. Chapter 8 described that culture from the inside. Here are its instruments, domesticated for a civilian with half an hour a day.

Upsolve

The problems you fail are the syllabus. The culture's single most agreed-upon practice: after any session, go back to what beat you — the same day, before the sting fades — and solve it properly. A problem is owned when it has been re-solved from a blank page, not when its solution has been read and understood. Understanding someone else's solution is a spectator sport ♦. Competitors say a problem has "entered the hands" only after the re-solve, and the phrase is exact: the hands, not the head.

The mistake journal

The non-negotiable instrument. Every miss becomes an entry, and every entry **names the move that was missed**, using the index's vocabulary. Five lines:

1. Date, and the puzzle (by chapter-number).
2. Where I stalled — one sentence.
3. The move I missed — *by name*, from Appendix A.
4. The recognition signal I ignored — one sentence.
5. The class of miss: *misread · wrong frame · right move, too late · unproven leap · bookkeeping*.

The classification line looks bureaucratic and is the most valuable line. "I am bad at puzzles" is a mood; "three of my last five misses were unproven leaps" is a diagnosis with a prescription attached — Chapter 10's exchange-argument discipline, applied until the leak closes. High-rated competitors keep exactly this ledger over years of contests; the journal is their error distribution made visible, and then made smaller.

Spaced review — of misses only

Re-solving your victories flatters; re-solving your failures teaches. The journal doubles as a review deck: each missed puzzle comes back at roughly three days, two weeks, and two months — re-solved cold, from the statement, with the re-solve time logged. Misses only. The wins need no rehearsal, and rehearsing them is how practice time quietly becomes entertainment ♦. This is spaced repetition, the most robust result in the learning

literature ∇ , pointed at reasoning patterns instead of vocabulary — a personalized curriculum in which every card earned its place by beating you.

The ladder: one star up

Comfort teaches nothing; panic teaches nothing. Competitive programmers phrase it as the +200 rule — practice about two hundred rating points above yourself. Here the unit is stars: work one star above the level you can solve cleanly. The signals are symmetrical. Weeks with an empty journal mean the ladder is too low; sessions that are all hints and no motion mean it is too high. Hard enough to hurt, close enough to finish.

Struggle before solutions

Thirty minutes of genuine motion before reading any solution — and *motion* has a definition: representations tried, cases enumerated, small instances computed. Rereading the statement and hoping is not motion. After thirty minutes, take one hint's worth — a sentence, a nudge, never the walk-through — and dive again. Reading solutions early trains a real skill: recognizing solutions. It is the wrong skill $\blacklozenge$. This is also why every solution in this book sits behind a pause line. The pause line is the pedagogy.

Check by a second road

Borrowed from engineers, who call it stress-testing and run it mechanically: when a contest solution looks right, they write a dumb brute-force twin and diff the two until a counterexample appears or confidence is earned. The civilian version costs two minutes: before declaring victory, ask *what independent computation would agree with me if I'm right?* Part III held itself to that standard — the circle-regions formula derived incrementally and then re-derived through Euler's formula; the twelve-coin scheme checked against its information budget. One road is an opinion. Two roads that meet are a result $\blacklozenge$.

Calibration: practicing "I don't know"

Chapter 13 argued that knowing-what-you-don't-know is among the most durable human edges — and showed what its absence costs in fabricated

citations and vaporized careers. The good news is that no moat in that chapter is cheaper to train. The instrument is a pencil.

The protocol: once a week, five questions. For each, build a Fermi chain — decompose, estimate, multiply, the move of 6:46 — and commit to a point estimate *plus an 80% interval*: "I am 80% sure the true number lies between these bounds." Then check, score the hits, and plot the running hit rate. The target is not narrow intervals. The target is honest ones: intervals labeled 80% that capture the truth 80% of the time. Widening your intervals until they stop lying is progress, however deflating it feels ♦.

Expect a rough start. In the classic experiments, people's "98% confident" intervals contained the truth roughly 60% of the time (Alpert and Raiffa's subjects, replicated for decades ♁). Overconfidence is the human default, not a character flaw — which is exactly why a logged curve beats a self-image. Question sources are everywhere: world numbers (a country's population, a river's length), personal numbers (unread emails, steps to the station), and this book's own numbers before you check them — how many arrangements of the 15 puzzle are reachable? Monty Hall belongs in this section for one reason: 6:33 is the canon's standing proof that confidence and correctness are different axes, even — especially — for experts. Ten thousand letters said so.

Two habits extend the log. First, tag everything: before checking any solution, write a percentage next to it. The habit generalizes far beyond puzzles, and it is the same habit this book's claim grades perform in print — ♁ versus ♦ is a confidence interval wearing typography ♁. Second, grade your sources the same way, including the machine ones. That becomes load-bearing two sections from now.

Insight hygiene

Insight resists scheduling, but it has hygiene — conditions under which the restructuring of Chapter 4 happens more often. Two are evidence-backed; one is a drill.

Incubation is real, and modest ♁. The meta-analysis of a hundred years of incubation experiments (Sio and Ormerod, 2009) finds genuine but small

effects, strongest when the impasse was honestly earned and the break is filled with something light — a walk, a chore — rather than more grinding or a second hard problem. The mechanism constraint matters ⚙: incubation consolidates work already done. Parking an unexamined problem parks nothing.

Sleep helps more ♾. In the *Nature* experiment that made the point stick (Wagner et al., 2004), solvers practiced a number task containing a hidden shortcut; after a night's sleep they were more than twice as likely to discover the shortcut as those who stayed awake the same interval. The protocol writes itself: struggle fully, write down the exact stuck-point in one sentence — the writing is the loading step — park it overnight, keep paper within reach. The shower is not magic. The loaded question is ⚙.

The drill: question the constraints. The trainable half of insight is not waiting for restructuring — it is hunting the restructure on paper. Stuck for ten minutes? Write every rule you are obeying. Mark each one STATED or ASSUMED. Attack the oldest ASSUMED first. This is 4·11's move run as a checklist, and 4·13's lesson made procedural: the boundary that blocks you was never in the statement. The object-form of the same drill comes from 4·14: for each thing in the problem, ask *what else is this, ignoring its label?* Three minutes, in writing, every time. Insight as checklist is less romantic than insight as lightning, and considerably more available ♦.

The sparring partner

As of July 2026, a frontier AI model will solve most of the classics in this canon on sight — they are in its training data — and will produce fluent, confident solutions to fresh puzzles that are usually right and occasionally, with identical confidence, wrong ♾. That combination makes it the best sparring partner ever built and the worst possible answer key. Three uses and one prohibition.

Generate variants. Ask for the same move in a new costume: "Here is a puzzle and the move it teaches [statement, plus the move's name]. Write three new puzzles that use the same move with entirely different surfaces — no ropes, no burning. Do not include solutions." This attacks surface-

binding, the failure mode the brain-training industry never solved, and it is Gick and Holyoak's compare-two-analogs condition, industrialized: the machine varies the costume, you extract the skeleton. Machines are excellent at costume.

Verify after committing. Solve first. Write your answer and your confidence percentage down — committed, dated, unfixable. *Then* ask the machine to attack it. If it agrees, note the hit. If it disagrees, one of you is wrong, and finding out which is a second workout — often better than the first, because now something is at stake. Puzzles are the safest ring there is for practicing this: unlike essays or strategies, a puzzle has ground truth, frequently checkable by enumeration. Part IV's forecast was that verifying confident machine output is becoming, durably, human work. Here you can rehearse that work with the answer key in the back of reality.

Coach the journal. Once a month, hand the machine your mistake journal and ask one question: "What pattern am I not seeing?" Summarizing a hundred small failures into three habits is precisely the kind of task it does well — and the diagnosis costs it no struggle you needed for yourself.

The prohibition: **never read its solution first.** Not as etiquette — as mechanism ⚙️. The struggle is the training stimulus; the solution is the anti-stimulus. Outsource the struggle and you keep the weakness, permanently, with excellent production values.

Sparring partner	Crutch
Variants before solving	Solutions before struggling
Attacks your committed answer	Answers instead of your attempt
One-sentence hint, on request, after 30 minutes	Walkthrough by default
Finds patterns in your journal	Replaces the journal

One symmetry deserves saying out loud: the machine is confidently wrong in exactly the way your calibration log is teaching you not to be. Keep a hit rate on its verdicts the way you keep one on your own intervals. Two calibration curves, one desk.

Train exactly what is durable

Chapter 13 mapped the moats. Every instrument in this chapter was chosen to sit on one:

The moat (Ch 13)	The training (this chapter)
Knowing you don't know	The calibration log; a confidence tag on every answer
Continual learning	The journal plus spaced misses — you update daily; models, as of July 2026, mostly update between versions 𐄂𐄂
Learning efficiency	Upsolving — maximum extraction from a sample size of one failure
Taste, and choosing what matters	The ladder you pick, the setter's mind of Ch 8, what you keep in maintenance
Ownership and accountability	Your name on the journal; your percentage on every answer

Nothing here needs an institution, a subscription, or a wearable. It needs twenty to forty minutes a day, a notebook, and a tolerance for being visibly wrong on paper — the one price fluency never asks and learning always does.

Chapter 16 is this chapter with dates on it.

CHAPTER 16

The Thirty-Day Regimen

Chapter 15 gave the instruments. This chapter puts dates on them:

C thirty days, twenty to forty minutes each, everything drawn from material already in this book. No equipment beyond a notebook (the journal), a timer, and occasionally coins, matches, or a deck of cards.

How to read the references. Puzzles are named by chapter·number: **5·6** means Chapter 5, canon puzzle 6; **C·30** means Appendix C — the Reserve Rack — canon puzzle 30, same ritual with no chapter warming you up. The number after the dot is the puzzle's canon number, printed in its heading where you first met it. Nothing is restated here — every reference lands on a self-contained ritual with the statement, the pause line, and the solution below it. **Cold** means: work from the statement only, with the solution covered. The spoiler discipline from *How to Use This Book* is now load-bearing.

You have seen these puzzles. That is the point, not a flaw in the plan. The regi-

men does not test whether you remember answers; it tests whether the *move* comes to hand — whether you can rebuild the path, name the pattern, and feel the recognition signal fire before the memory of the answer arrives. Re-solving from a blank page is a different act from recalling, and the journal (which demands the reasoning in writing, not the answer) keeps the two honest. Where memory does contaminate, the clock catches it: rebuilding is fast when the move is owned, halting when only the answer is.

The rules, once:

1. Twenty to forty minutes. Set the timer. Stop at forty — the edge, not exhaustion.
2. The journal is half the training. Every miss gets the five-line entry from Chapter 15, and the missed move gets *named*.

3. Struggle rule: twenty minutes of real motion before any peek; then one hint's worth, then back in.
4. A miss is upsolved the same day or it enters the review deck. Nothing is abandoned.
5. A skipped day is skipped. Do not double up; the deck keeps score, not the calendar.
6. AI only per Chapter 15: variants and post-commitment attacks. Never its solution first.

Deck day (each week's last day) means: re-solve, cold, every miss the journal has collected, and log the re-solve times. **Calibration set** means: five Fermi questions with 80% intervals, scored, per Chapter 15's protocol.

Day zero, ten minutes. Copy the five-line entry template onto the journal's inside cover. Bookmark Appendix A; it will be consulted daily. Then run calibration set 0 — five questions, five intervals, scored and dated — so the curve has a baseline nobody can later flatter. The regimen measures change, and change needs a before.

The anatomy of a session

The same thirty minutes, dissected, so the schedule below can stay terse:

- **Minutes 0–3 — read and frame.** Read the statement twice. Write one line: what kind of problem this *smells* like, and which move the smell suggests. This line is the day's real bet; the journal will grade it.
- **Minutes 3–23 — motion.** Representations tried, cases enumerated, small instances computed. When stuck for ten of these minutes, run Chapter 15's constraint drill: every rule you are obeying, marked STATED or ASSUMED.
- **Minute 23 — the hint decision.** Still nothing? One hint's worth: reread the puzzle's Move line only — the bolded name, not the paragraph — then back in.
- **Last 5 minutes — the ledger.** Solved or not, write the entry: the move, named; the signal, in a sentence; the miss class if it was a miss. On a miss, schedule the upsolve for today or file the card.

Two people can run the regimen on one book, and it improves: alternate days as solver and verifier, with the verifier holding Chapter 15's rules — one-sentence hints on request, attack after commitment, never the solution. Explaining a move you just used is the cheapest known way to find out whether you own it ♦.

Week 1 — Frames and deduction

The insight muscles of Chapter 4, then the propagation discipline of Chapter 5.

Day	The work (20–40 min)	Into the journal
1	4-11, 4-12, 4-13 cold. Before each: write, in one line, the frame you think you are in.	First entries. Name each move from Appendix A, hit or miss.
2	4-14, 4-15. Then the drill: pick three household objects; for each, write what else it is, ignoring its label.	Which assumption you held longest, and who put it there.
3	4-16, then 5-2. Finish by inventing one knights-and-knaves exchange of your own that has a unique solution.	The recognition signal for each move, one sentence apiece.
4	5-3, 5-4.	Where a guaranteed error became usable information.
5	5-5 — the full grid, one sitting. A 40-minute day.	Count your guesses. Were any made while a forced move was still on the table?
6	5-6, then 5-7.	The meta-fact you almost wasted; the greedy you almost trusted.
7	Deck day. Every Week-1 miss, cold, timed.	Naming rate: misses named ÷ misses logged.

Week 2 — Numbers and chance

Chapter 6: conservation, recursion, digits under constraint, and the probability moves that punish good intuition.

Day	The work (20–40 min)	Into the journal
8	6·26, C·29.	Which conservation ended which argument.
9	C·30, C·31. Play C·31 with three coins on three saucers until the recursion is in your fingers.	The pairing, and its exception. The smaller self you trusted.
10	6·41, and 6·48 if time remains.	The tightest constraint — did you start there, or wander in?
11	C·21, C·47.	Where the gut objected; what the algebra said back.
12	6·32, then 6·33. Then deal three cards and run twenty rounds of 6·33 both ways, tallying.	Your simulated frequency next to the theoretical one.
13	6·34, 6·35.	One sentence each: what procedure generated the data.
14	6·46, then calibration set 1. Deck day for the week's misses.	Interval hits: n of 5. First point on the curve.

Week 3 — The olympiad moves

Chapters 8–11. The format shifts: each day, re-derive a worked example cold — statement only, path from scratch — then run its drill. These are the sixteen tools' home gyms.

Day	The work (20–40 min)	Into the journal
15	Pigeonhole (Ch 9): the five-points-in-a-square proof, cold. Then the drill pair: same-hair-count Londoners; the party of equal friend-counts.	The boxes you invented — and what forced their number.

Day	The work (20–40 min)	Into the journal
16	Invariants (Ch 9): the 1-to-10 blackboard reduction, cold; the chameleons drill. Then revisit 2-38's swap as a story about dynamics.	The quantity that survived the move. How you found it.
17	Extremal and symmetry (Ch 9): the tournament king; the round-table game. Then play C-43 with heaps of 3, 5, 7 coins until the balancing reply is a reflex.	The extremal candidates you enumerated before one bit.
18	Coloring (Ch 9): the 10×10 board against 1×4 tiles, cold. Then the T-tetromino drill.	The first coloring you tried, and why the tile's shape vetoed it.
19	Induction and backwards (Ch 10): the any-cell tromino tiling step; then the race to 100 against an opponent (or your left hand).	Where the plain hypothesis leaked, and the cargo you loaded into it.
20	Modeling and double counting (Ch 11): re-derive Königsberg from a blank map; the classroom census; then a full 40-minute attempt at the fixed-points identity (IMO 1987, Problem 1) under the struggle rule.	The incidence set, named explicitly — "pairs (x, y) such that..."
21	Deck day , plus calibration set 2 .	Naming rate; interval hits. Two points make a curve.

Week 4 — The boss room, and the mirror

Chapter 7's monsters, one per day — a single ★★★★★ is a full session — braided with the calibration thread. The week ends by measuring you.

Day	The work (20–40 min)	Into the journal
22	7-8. Before touching a pan: write the information budget — how many answers, how many outcomes per test.	Did the budget precede the design? The move, named.
23		

Day	The work (20–40 min)	Into the journal
	7-9. Design your own protocol on paper before reading anything. Then compare.	Where your protocol leaked — the double-count, the ambiguity.
24	7-10. Walk the induction out loud for one, two, three islanders before scaling it.	The rung where it stopped feeling like a trick.
25	7-37. Then shuffle a deck and run ten trials of the strategy, following the cycles by hand.	Simulated survival rate beside the true one; the correlation sentence, in your own words.
26	7-44, then calibration set 3 .	The endgame you anchored first. Interval hits.
27	7-49 — the summit. Full forty minutes; the only hint permitted is rereading 5-6.	Every filter in the dialogue, restated in your own words.
28	7-50. Write one page: where you think the flaw lives, and your confidence in percent. Then read the candidate diagnoses again.	Your diagnosis and percentage — filed where you can be embarrassed by it later, which is the practice.
29	The audit. No new puzzles. Compute: naming rate across four weeks; the calibration curve (fifteen intervals by now); re-solve times, first attempt versus deck. Write your top three miss classes.	The three numbers, one sentence each. Your personal syllabus, in your handwriting.
30	The exit. (1) Re-solve your five stubbornest misses, cold. (2) Build one puzzle of your own around your favorite move — kernel first, costume second, per Chapter 8's setter's craft — and give it to a person. (3) Write next month's maintenance plan.	The puzzle you built and the move it hides. Whether your solver found it.

Why this order

The sequence is engineered, not decorative. Frames come first because every later week depends on noticing self-imposed constraints, and that noticing is fastest to install (a week of ★1–2 puzzles buys it). Deduction follows immediately so that the propagation discipline — no guessing while forced moves remain — hardens before the probability material arrives to punish sloppy conditioning. Numbers and chance sit in week 2 because they supply the regimen's first calibration data and its first honest humiliations; Monty Hall works better on a solver who has already kept a journal for eleven days. The olympiad moves occupy week 3 as deliberate cross-training: chapter-sized tools, re-derived rather than re-read, precisely when the first two weeks' moves risk settling into comfort. And the boss room is last not as a graduation ceremony but because ★★★★★ puzzles are *composites* — 7·8 is an information budget wearing a deduction costume; 7·37 is probability riding structure — and composites are only trainable once their parts are reflexes. The mirror days (29–30) close the loop the journal opened on day 1: the regimen ends by measuring itself, which is the habit it most hopes to leave behind.

Maintenance mode

After day 30, two sessions a week hold everything the month built. Fewer, and the recognition signals fade to trivia; more is optional, not required.

Session A — new material (30 min). One of: a canon puzzle the regimen never scheduled (the bench: 6·22, 6·48 if either went unused, and the rest of Appendix C — C·23, C·24, C·36, C·45 among its seventeen); a Part III ★★★★★ attempted for real under the struggle rule; or one problem from the outside world — the doors listed under Further Reading (a Gardner collection, Project Euler, an ARC grid) — given the full ritual: struggle, solve, *name the move*, journal the miss.

Session B — the deck (20 min). Every miss aged past two weeks, re-solved cold, re-timed. Monthly: one calibration set, and the journal audit — naming rate, curve, top miss class. If a class refuses to shrink, its chapter becomes Session A material until it does. The prescriptions:

Recurring miss class	The prescription
Wrong frame	Chapter 4 bench work; the constraint drill before every session for two weeks
Guessed while forced moves remained	Chapter 5 re-solves with a guess-counter in the margin
Conditioning errors, procedure ignored	Chapter 6's probability block; simulate with cards until theory and tally agree
Unproven leap	Chapter 10's exchange discipline; no greedy claimed without its one-line proof or its counterexample
Bookkeeping	Slow down 20%; check by a second road before declaring — Chapter 15's sixth instrument

The standing rule. Any puzzle met in the wild — interview, pub, timeline — gets the ritual. The regimen ends; the ritual doesn't.

Honest progress measures

Three numbers, none of which can be gamed without noticing.

1 • Misses named. The fraction of journal misses where the missed move got its name from Appendix A. Expect something near half in week 1 and something near all by week 4 — not because you miss less (the ladder should prevent that) but because the vocabulary has become perception. This number is the skill: recognition, measured at the moment of failure.

2 • The calibration curve. Your 80% intervals should capture the truth about 80% of the time — after fifteen or twenty intervals, a hit rate between 70 and 90 is honest work. Beware the failure that looks like success: intervals so wide they always hit train nothing. Precision under honesty is the game — intervals that narrow *while* the hit rate holds.

3 • Re-solve time on old misses. The deck's clock, falling. A miss that took thirty minutes and three hints re-solved in six minutes flat has entered the hands. A miss whose re-solve stalls at twenty minutes twice running is telling you which chapter to reread — believe it.

Expect the numbers to misbehave, and know what the misbehavior means. A **plateau** in naming rate around week 3 is normal: the olympiad tools temporarily flood the vocabulary, and perception lags acquisition by a few days. A **falling** calibration hit rate is usually good news wearing bad news's coat — it appears exactly when intervals start narrowing, and the fix is to narrow slightly less, not to stop narrowing. **Regression** on the deck — an old miss missed again — is the single most informative event the regimen produces: it marks a move that was memorized rather than owned, and it earns an immediate second journal entry with a different answer on line 4 than last time. If the same puzzle produces the same named miss twice, the name is wrong; the real gap is one level deeper, and a session staring at that gap is worth a week of new material.

And the non-measures, named so they can be refused: streak length, puzzles per day, hours logged, and how smart a session felt — the last being fluency, Chapter 15's documented liar ♁. Thirty days buys no measurable rise in anything an IQ test totals, and this book will not pretend otherwise. It buys thirty-two moves at recognition speed, a journal that knows your failure modes better than any test, and a calibration baseline most professionals never acquire. That is not everything. It is more than almost anyone has.

The gate

Thirty days is not mastery. It is something better: working habits, and a vocabulary your eyes now run without being asked.

The Sphinx has been waiting at this book's gate since Chapter 1, and she has not aged. She was never really a monster on the road to Thebes. She is every gate with a question on it that punishes the literal reading: the interview problem built to be misread, the too-clean chart, the contract clause whose surface hides its instruction, the confident machine output that is fluent, cited, and wrong. Solve or be eaten — updated terms, same gate.

Her first riddle — four legs at morning, two at noon, three at evening — is, you may notice, still open. Sixteen chapters, and this book has not answered it. Books must close their boxes, and this one closes it on its final page,

where the colophon keeps the answer. Go and look whenever you like. But mark what these thirty days changed: you no longer need the answer to get past her. You have the move. Answers fit one lock apiece; a move opens a corridor of them — which is why the index at the back is the only part of this book built to be reread for the rest of your life.

The manual ends. The gymnasium doesn't.

Appendix C — The Reserve Rack

The chapters used thirty-three of the canon's fifty puzzles; these seventeen complete the set. No chapter warms you up here, and no section title whis-pers the move in advance — which is exactly their value to the Chapter 16 regimen. Fresh meat: solve them cold.

★ 17 • The Riddle of the Sphinx

"What goes on four legs in the morning, two legs at noon, and three legs in the evening?"

The Sphinx is waiting.

Solution. Withheld — the one puzzle in this book whose answer you must earn by finishing it. It has stood at the gate since Chapter 1, and the colophon settles it on the book's final page. If the wait is unbearable, that is the point of the wait.

The Move — Change the timescale. When a literal reading yields a monster, re-scale time or re-bind a word — every noun in a riddle is a metaphor slot.

★★ 21 • The Rope Around the Earth

A rope hugs the Earth's equator (~40,000 km). You splice in one extra meter and float the rope uniformly above the surface. Can a cat slip under? And if instead you pull all the slack taut upward at one point, how high is the peak?

Commit to a guess — cat or no cat — before reading on.

Solution. The gap is the extra circumference divided by 2π : $1/(2\pi) \approx 15.9$ cm, all the way around. The cat strolls under. Better: the answer is independent of the planet's size — $C = 2\pi r$, so $\Delta r = \Delta C/2\pi$, and the radius cancels; a

marble gives the same 15.9 cm. The second question is just as rude (verified numerically): pulled taut at one point, the peak sits about 121 m above the surface ($\theta \approx (3s/2R)^{1/3}$) with $s = 1$ m of slack; height $\approx R\theta^2/2$). Intuition fails in both directions.

The Move — Let the algebra veto the gut. Feelings scale with absolute size and the relation $\Delta r = \Delta C/2\pi$ doesn't; when a two-line derivation and intuition disagree, the derivation wins.

★★★ 23 • The Coin Rotation Paradox

Coin B rolls without slipping once around an identical, fixed coin A. How many times does B rotate about its own axis? Then: a circle of radius r rolls around a circle of radius $3r$ — how many rotations?

Get two coins and try it. Really.

Solution. Two, not one — roll a coin around another and watch: the head is already upright after half the journey. Rotation = rolling along the path + the orbit itself: contact-rolling contributes $C_A/C_B = 1$, and the orbit contributes 1 more. The general exterior formula is $R/r + 1$, so the radius-3 case gives 4, not 3. That exact question appeared on the May 1982 US SAT with "3" keyed as correct; three students proved the test wrong, ETS conceded, and every score was adjusted. The sky runs the same phenomenon: the sidereal versus the solar day — Earth rotates 366.25 times in a year of 365.25 days.

The Move — Separate spin from orbit. When motions compose, decompose them — and when a count is disputed, count a second, independent way and demand agreement.

★★★★ 24 • The Spider and the Fly

A room 30 ft long, 12 ft wide, 12 ft high. A spider sits on one 12×12 end wall, centered, 1 ft below the ceiling. A fly sits on the opposite end wall, centered, 1 ft above the floor. What is the spider's shortest walking path to the fly, along any surfaces?

Sketch the room before you trust yourself.

Solution. The obvious route — straight up, along the ceiling, straight down — is $1 + 30 + 11 = 42$ ft. But distances on a box are straight lines on its unfoldings. Flatten the right five-face net — end wall, ceiling, side wall, floor, far end wall — and the path becomes the hypotenuse of a 32×24 right triangle: $\sqrt{(1024 + 576)} = \sqrt{1600} = 40$ ft. (Intermediate unfoldings give 43.2 and 40.7; 40 is best, machine-checked.) A geodesic touching five of the six faces beats the route that looks direct. Dudeney, 1903 — Gardner called it one of the best puzzles ever devised.

The Move — Unfold the space. Change the representation until the answer becomes a straight line: reflect the billiard table, unroll the cylinder, take logs.

★★★★ 25 • The Haberdasher's Problem

Cut an equilateral triangle into as few pieces as possible that reassemble into a perfect square.

Scissors down. Think first.

Solution. Four pieces — and, astonishingly, they can be hinged in a chain that swings one way into the triangle and the other way into the square. Construction sketch: bisect two sides at D and E; place F on the third side so that EF equals the side of the target square ($\sqrt{\text{area}}$); perpendiculars complete the three cuts — the full compass-work is in Dudeney (1907). Existence was never in doubt: any polygon dissects into any equal-area polygon (Wallace-Bolyai-Gerwien theorem, 1807–1833); the game is minimality and eleg-

ance. Dudeney posed the problem on 6 April 1902, had the four-piece answer by 4 May, and showed a brass-hinged model to the Royal Society in 1905. In December 2024, Demaine, Kamata and Uehara proved that no three-piece dissection exists (arXiv 2412.03865) — the 122-year-old answer, finally certified optimal.

The Move — Re-partition the same area. Conservation guarantees a bridge between equal-area shapes and craft is finding the minimal one — proving you can't do better can be a century harder than doing it.

★ 27 • Gauss's Schoolroom Sum

Sum the integers from 1 to 100 — fast, in your head.

You have ten seconds, the story says.

Solution. Pair the ends: $1+100$, $2+99$, ..., $50+51$ — fifty pairs, each summing to 101: **5050**. In general, $n(n+1)/2$ — the sum counted forward and backward at once is n copies of $(n+1)$. About the legend: the schoolboy-Gauss anecdote (slate slapped down, "ligget se") is real in outline but embellished in detail — the earliest source (Sartorius, 1856) names no specific series, and the "1 to 100" specifics accreted later (Hayes, 2006). The method is ancient regardless.

The Move — Pair the extremes. Impose a symmetry the data doesn't advertise, and the computation collapses.

★★ 28 • The Fly Between the Trains

Two trains 20 miles apart approach each other, each at 10 mph. A fly leaves one train at 15 mph, flies to the other, turns instantly, flies back, and so on until the trains meet. How far does the fly fly?

Resist the urge to sum a series.

Solution. The trains close at 20 mph, so they meet in exactly 1 hour. The fly flies at 15 mph for that hour: **15 miles**. No series needed — though the infinite zig-zag sum converges to the same answer, which is the joke in the (probably polished) von Neumann anecdote. Told the puzzle, he answered instantly. "Ah, you knew the trick." — "What trick? I summed the series."

The Move — Total the aggregate, skip the episodes. When a process is a mess of episodes but a global quantity — time, distance, energy — is clean, compute the global.

★★ 29 • Wine and Water

A glass of wine and a glass of water, same volume. Move one spoonful of wine into the water, stir (or don't — it doesn't matter), then move one spoonful of the mixture back. Is there more wine in the water, or more water in the wine?

Take a stance before the argument starts.

Solution. Exactly equal — regardless of stirring, spoon size, or how many round trips are made. Endstate argument: both glasses finish at their original volumes, so whatever volume of wine is missing from the wine glass must be occupying the water glass, and the space it left behind is filled by exactly that much water. Swap-symmetry settles the matter before any tracking of concentrations begins — the chessboard coloring's skeleton, in liquid form (a Rouse Ball and Gardner staple).

The Move — Conservation ends the argument. When the bookkeeping looks hairy, ask what the final totals must be and let conservation do all the arithmetic.

★★★ 30 • The Thousand Lockers

Lockers 1–1000, all closed. Student k ($k = 1 \dots 1000$) toggles every k th locker. After all 1000 students pass, which lockers stand open?

Try lockers 1 through 10 by hand first.

Solution (verified by simulation). Locker n is toggled once per divisor of n , so it ends open iff n has an odd number of divisors. Divisors pair up, $d \leftrightarrow n/d$ — except when $d = n/d$, that is, when n is a perfect square. The open lockers are 1, 4, 9, ..., 961: **the 31 perfect squares ≤ 1000** . The exception to the pairing is the answer.

The Move — Pair things up; interrogate the exceptions. For parity questions, find the involution and look hard at its fixed points.

★★ 31 • The Tower of Hanoi

Three pegs; n disks stacked by size on one peg. Move the whole tower to another peg, one disk at a time, never placing a larger disk on a smaller. Minimum moves? (The monks of the legend run $n = 64$.)

Three disks by hand teach more than an hour of staring.

Solution. $T(n) = 2T(n-1) + 1$: move $n-1$ disks aside, move the big disk, restack — so $T(n) = 2^n - 1$, and the recursion is the algorithm. For the monks' 64 disks: $2^{64} - 1 = 18,446,744,073,709,551,615$ moves — at one per second, about 585 billion years (the wheat-on-the-chessboard number; doubling incarnate). The "ancient Brahmin legend" was invented wholesale by the puzzle's actual author, Édouard Lucas, who sold it in 1883 under the anagram-pseudonym "N. Claus de Siam" — marketing and mathematics in one box.

The Move — Recurse: trust the smaller self. Solve the problem assuming a solved smaller instance and let the structure carry you — the puzzle most worth feeling in the hands before ever writing a recursive function.

★★★★ 36 • The Two Envelopes

Two sealed envelopes, one holding twice the other's amount. You pick one — it holds X . The argument: "The other holds $2X$ or $X/2$ with equal chance; expected value $(2X + X/2)/2 = 1.25X > X$, so switch. But the same argument says switch back, forever." Where exactly does it break?

Find the precise sentence where the argument cheats.

Solution. The flaw is a variable shift. In the branch "other = $2X$," the symbol X denotes the smaller amount; in "other = $X/2$," it denotes the larger. One letter, two meanings — the formula adds apples to oranges. Done properly: call the pair $(Y, 2Y)$; whichever envelope is in hand, switching gains $+Y$ or $-Y$ with equal probability — expected gain **0**. The seductive version silently assumes all amounts equally likely, a uniform prior on an infinite range, which is not a probability distribution; with any proper prior, the conditional expected gain integrates to zero. (For a specific observed amount, a proper prior can genuinely favor switching — that is Bayesian updating, not paradox.) Ancestry: Kraitichik's necktie paradox (1943); the modern treatment is Nalebuff (1989).

The Move — Pin the random variable. Write down what is random, what is observed, and what each symbol denotes in each branch — most probability "paradoxes" are notation crimes.

★★★★ 39 • Rubik's Cube and the Commutator

Not "solve the cube" — the design problem underneath it. Find move sequences that change only a little — cycle three corners, say — while leaving the rest of the cube's 4.3×10^{19} -state chaos untouched. How does one even search for such a needle?

This one asks for a method, not an answer.

Solution. The commutator: for sequences A and B, perform $A B A^{-1} B^{-1}$. If A and B touch mostly disjoint pieces, almost everything cancels, and the net effect is a small local change — a three-cycle of corners or edges, a pair of twists. Conjugation ($C X C^{-1}$) then aims the tool: carry the pieces you care about into position, apply the commutator, carry them back. Every human method is a library of conjugated commutators; that is why beginners' "algorithms" exist at all. The object's scale: 43,252,003,274,489,856,000 positions — yet God's number is **20** moves (Rokicki-Kociemba-Davidson-Dethridge, 2010, ~35 CPU-years), while human methods spend 50–60. The gap between what exists and what is findable, held in the palm.

The Move — Build commuting tools. Do, do, undo, undo: engineer operations whose interference is exactly the change you want, then conjugate to aim them — how one edits any system too complex to rebuild, codebases and organizations included.

★★ 40 • The Caesar Shift and the Frequency Attack

Decrypt: **YMJ RFU NX YMJ YWJFXZWJ** (a Caesar cipher — every letter shifted by the same amount).

Count letters before you count keys.

Solution. Only 25 keys exist, so brute force is instant — but the elegant attack is statistical. The ciphertext's most frequent letter is J (four times); English's most frequent is E; the hypothesis $J = E$ means shift 5. Apply shift -5: **THE MAP IS THE TREASURE**. The same idea scales to general substitution ciphers, where $26!$ keys kill brute force: match letter-frequency fingerprints (E T A O I N ...), then common digrams (TH, HE), and the cipher unzips. That is al-Kindi's 9th-century insight — statistics defeating secrecy a millennium before "statistics" was a word.

The Move — Attack the statistics. Encryption hides content, not structure; whatever must preserve function must preserve pattern, so find the pattern that couldn't be hidden.

★★★ 42 • Two Cryptic Clues

Solve as cryptic-crossword clues: (a) "**Gegs (9,4)**" (b) "**HIJKLMNO (5)**".

Say each clue aloud. Twice.

Solution. (a) **SCRAMBLED EGGS** — the clue demonstrates its answer: "gegs" is literally eggs, scrambled; the enumeration (9,4) confirms. (b) **WATER** — the letters run from H to O: "H-to-O" = H₂O. Both are definition-by-performance: the clue doesn't describe the answer, it enacts the instruction that produces it. Beneath them sits standard cryptic grammar: every fair clue contains a definition and a wordplay path — two independent proofs of one answer, the Ximenean contract (Ximenes, *On the Art of the Crossword*, 1966).

The Move — The surface hides an instruction. Read strange text twice — once as description, once as recipe — because sometimes the form, not the content, is the payload.

★★★ 43 • Nim

Heaps of counters — say 3, 5, 7. Players alternate removing any positive number of counters from one heap. Taking the last counter wins. Who wins, and how? (Warm-up: from 21 flags, players alternately remove 1–3; taking the last wins — who?)

Play the 21-flag warm-up against yourself first.

Solution (theorem machine-verified for all heaps $\leq 5,5,5$). Chapter 9 derives this balance as a toolkit weapon; here is the pocket edition. Warm-up: multiples of 4 are losing positions — leave 20, 16, 12, 8, 4 and mirror. Full Nim: write the heap sizes in binary and XOR them — the **nim-sum**. Position (3,5,7): $011 \oplus 101 \oplus 111 = 001 \neq 0$, so the first player wins; the winning reply is $3 \rightarrow 2$, making $010 \oplus 101 \oplus 111 = 0$. Strategy: always move to nim-sum 0. From 0, every move breaks the balance — the opponent must hand back a nonzero

position — and from nonzero, a balancing move always exists. Bouton proved this completely in 1901, the first game ever solved; Sprague and Grundy (1935–39) then showed every impartial game is secretly Nim — one puzzle's invariant swallowed a genre.

The Move — Hold the balanced state. Identify the invariant class of dead positions and return the game to it every turn, until the opponent's freedom is irrelevant.

★★ 45 • Guess Two-Thirds of the Average

Everyone in a group privately picks a number in $[0,100]$. Closest to two-thirds of the group average wins. What do you pick?

Pick your number before you meet the theory.

Solution. Pure theory: any guess above 66.7 is dominated, since the average cannot exceed 100. If nobody plays dominated strategies, the average cannot exceed 66.7, so guesses above 44.4 die next; the elimination iterates all the way down to **0**, the unique Nash equilibrium. Reality: in Nagel's 1995 experiments, and in newspaper contests with thousands of players, the winning number sits around **20–25** — real populations are a mixture of level-0 players (random, ~50), level-1 (guess 33), level-2 (guess 22), with few going deeper. The right play is not the equilibrium; it is one step deeper than the crowd, no more. Keynes' beauty-contest description of markets (1936) is this puzzle wearing a suit.

The Move — Iterate other minds (then stop). Model what others will do, including their model of you, but calibrate the depth of recursion to the actual population rather than an ideal one.

★ 47 • Fold Paper to the Moon

A sheet of paper is 0.1 mm thick. Fold it in half 42 times (grant the physical impossibility). How thick is the stack — a meter? A building? More?

A meter, a building, or more — commit.

Solution (computed). $2^{42} \times 0.1 \text{ mm} \approx 4.4 \times 10^8 \text{ m}$ — **past the Moon** (mean distance $3.84 \times 10^8 \text{ m}$). Twenty folds already exceed 100 m; thirty reach about 107 km. Physical reality check: the world record is **12 folds** — Britney Gallivan, 2002, with a 1.2 km roll of tissue, after first deriving the loss function $L = (\pi t/6)(2^{n+4})(2^n - 1)$, which shows why each fold costs quadratically more paper. Sibling legend: the wheat-doubling chessboard, $2^{64} - 1 \approx 1.8 \times 10^{19}$ grains (told of Sissa by Ibn Khallikan, 1256).

The Move — Respect the exponent. Intuition extrapolates linearly and doubling punishes that within thirty steps: when a process multiplies — debt, virality, compute, compounding error — calculate, never eyeball.

Appendix A — The Moves Index

Fifty puzzles and sixteen olympiad tools compress to thirty-two working moves. One line each; the line is a reminder, not a lesson — the lesson lives at the reference. References read chapter·puzzle: **5·6** is Chapter 5, canon puzzle 6; a plain **Ch 9** points to a chapter's worked material rather than a numbered puzzle; **C·17** means Appendix C — the Reserve Rack — canon puzzle 17. This index is the book's real curriculum. The puzzles were the delivery system.

See it differently

- **Switch the frame; question the boundary.** When the facts are bizarre but consistent, the blocking constraint is usually one you added — 4·12, 4·13, C·17.
- **Break functional fixedness.** Ask what each object is, ignoring its label: a box is also a shelf — 4·14.
- **Change the space.** Add a dimension when the plane is the prison (4·15); unfold the space until the answer is a straight line (C·24).
- **Model the story as structure.** Name the vertex and the tale collapses to a known kernel; measure progress in state-space, not cargo — 1·1, Ch 11.

Find what cannot change

- **Find the invariant.** A conserved quantity refutes without search — colorings for tilings, parity for move-systems; or find the monovariant that only falls, and the process must stop — 6·20, 2·38, Ch 9.
- **Conservation ends the argument.** Endstate totals and global aggregates beat episode-by-episode tracking; when positions lie, rates may not — C·29, C·28, 4·16.
- **Hold the balanced state.** Identify the dead positions and return the game to them every turn; the opponent's freedom becomes irrelevant — Ch 9, C·43.

Treat knowledge as a substance

- **Budget outcomes before designing tests.** Count the possible answers, count each test's outcomes, and let the ratio dictate the design — 7·8, Ch 11.
- **Assume, then break it — and use the guaranteed error.** Hypothesize each case and let contradiction do the elimination; a certain lie is exactly as informative as a certain truth, and can be routed to cancel itself — 5·2, 5·3, 5·4.
- **Declared ignorance is data.** "I don't know," said aloud, filters worlds; public words mint common knowledge, which is more than everyone knowing — 5·6, 7·49, 7·10.
- **Condition on the procedure.** What a fact means depends on how it reached you; the sampling defines the odds — 6·33, 6·34.
- **Knowledge can self-destruct.** Statements about what you will know are not ordinary facts; feeding them into your reasoning can invalidate them — 7·50.

Compute a smarter way

- **Anchor, then propagate.** Plant the absolute facts, squeeze the tightest constraint first, and never guess while a forced move remains — 5·5, 6·41.
- **Work backwards.** The goal end of the tree is thinner; label the endgame and roll toward the present — 7·44, Ch 10.
- **Recurse, then cache.** Trust the smaller self; the art is choosing what a "state" must remember — C·31, Ch 10.
- **Count another way.** Complement (6·32), pairing and its exceptions (C·30), double census, expectation — one set, two ledgers, and the agreement is your theorem; an average of v forces some case $\geq v$ — Ch 11.
- **Small cases propose; proofs dispose.** Tabulate, conjecture, then prove — and when the inductive step leaks, strengthen the claim until it carries its own fuel — Ch 10, Ch 11.
- **Prove greed with an exchange.** Or spend the same ten minutes hunting its counterexample; unproven greed is a leap, not a plan — Ch 10.

- **Binary-search the answer.** When feasibility is monotone, optimize by deciding, log-many times — Ch 10.
- **Interrogate the extreme witness.** The largest, smallest, closest, or first element arrives pre-loaded with contradictions for free — Ch 9.
- **Impose or steal a symmetry.** Mirror the opponent and every future question is pre-answered; or show the extra move never hurts — Ch 9.
- **Invent the pigeonhole boxes.** The theorem is one line; the entire craft is manufacturing the categories — Ch 9.
- **Declare WLOG, then design the casework.** Quotient the symmetry first; choose the split so every branch lands on a finishing move — Ch 11.

Borrow, share, correlate

- **Restructure the bill.** Borrow a unit and return it (6·26); make the expensive items share (5·7). Temporary extensions reveal permanent structure.
- **Engineer the joint outcome.** Correlate fates so everyone lives or dies on one event (7·37); when shared state is tiny, design the protocol, not the deduction (7·9).
- **Build commuting tools.** Do, do, undo, undo — small operators that change little and preserve the rest are how you edit any system too complex to rebuild — C·39.

Audit before trusting

- **Audit the invisible assumption.** List your defaults in writing; attack the oldest first — 4·11.
- **Distrust the surface.** Verify the diagram, run the algebra against the gut, attack the statistics; whatever must preserve function must leak pattern — 6·22, C·21, C·40.
- **Pin the random variable; stratify before comparing.** Most paradoxes are notation crimes; most reversals are hiding in the case mix — C·36, 6·35.

Scale

- **Decompose, estimate, multiply — then respect the exponent.** Work in powers of ten, state bands not points, and never eyeball a doubling process — 6·46, C·47.

Other minds

- **Model the maker.** Go one level deeper than the crowd, then stop (C·45); artifacts built by intent are decoded fastest by modeling the intent (Ch 8) — and sometimes the surface is the instruction (C·42).
- **Seek the fixed point.** Make the description part of the described and demand consistency: $x = f(x)$ — 6·48.

Appendix B — The Graveyard of Thinking Myths

Every discipline keeps a trophy room; this book also keeps a graveyard. The six residents below were beautiful, plausible, and beloved. Several were founded on real experiments. All are dead. They are buried with respect and full honors, because each taught the field something by dying — and because a book that spends fifteen chapters teaching skepticism owes its readers the kill list. Visit; leave flowers; do not resurrect.

⊃ Commercial brain training

The claim. Ten minutes of app puzzles a day raises memory, attention, and intelligence, the way curls raise a bicep. **Why it spread.** Real neuroplasticity science, an aging population happy to pay for insurance against decline, and honest-feeling receipts: scores on the trained games genuinely climb. **What the evidence says.** Owen et al. (*Nature*, 2010): 11,430 adults, six weeks — trained tasks improved sharply, transfer even to close cousins was approximately nil ∅. Simons et al.'s 2016 review of the industry's evidence: the same. The FTC fined Lumosity's maker \$2 million in 2016 for deceptive advertising. Practice is specific: the game trains the game. **Verdict.** ⊃ Dead. Epitaph: HE GOT REALLY GOOD AT THE GAME. (Chapter 15's naming-and-recostuming discipline exists because this corpse is real.)

⊃ Learning styles

The claim. Each person is a visual, auditory, or kinesthetic learner, and teaching matched to the style produces better learning. **Why it spread.** Preferences are real and easy to measure; the idea flatters individuality; an inventory-and-workshop industry grew around kind teachers who wanted to personalize. **What the evidence says.** Pashler, McDaniel, Rohrer, and Bjork (2008) audited the field: the claim requires a crossover interaction — visual learners learning *more* from visual teaching, auditory learners the reverse — and almost no study ran the right design; those that did found no meshing effect ∅. Preferences exist; matching to them doesn't help. The

content dictates the best medium: geometry wants pictures, for everyone. **Verdict.** ☞ Dead. Teach to the material's style, not the student's.

☞ **Left-brained and right-brained people**

The claim. Logical people run on the left hemisphere, creative people on the right. **Why it spread.** Sperry's split-brain research was real, spectacular, and Nobel-crowned (1981); personality quizzes and corporate workshops flattened a true fact about severed hemispheres into a false fact about accountants. **What the evidence says.** Functions lateralize; people don't. Nielsen et al. (2013) analyzed resting-state scans of 1,011 brains and found no individuals with a globally dominant hemisphere ☹. Every intact brain runs both halves on essentially everything. **Verdict.** ☞ Dead. You are whole-brained; so is everyone you have ever argued with.

☞ **We use only 10% of our brains**

The claim. Nine-tenths of your brain idles, waiting to be unlocked. **Why it spread.** Often traced to a mangling of William James, who wrote (1907) that we use a fraction of our mental *resources* — a motivational claim about effort; the tidy "10%" was pinned on later and no one traceable ever measured it. It survives because it prints permission slips: the missing 90% is yours, for the price of the seminar. **What the evidence says.** Imaging shows nearly all of the brain active across an ordinary day; no region can be lesioned for free ☹; and an organ that is 2% of body mass billing roughly 20% of the energy budget would be evolution's most overpaid idle employee ☹. **Verdict.** ☞ Dead. You use all of it. Some days it doesn't feel that way, which is a different problem.

☞ **The Mozart effect**

The claim. Listening to Mozart raises IQ — so play sonatas to infants and they grow up smarter. **Why it spread.** Rauscher, Shaw, and Ky (*Nature*, 1993) found that 36 college students improved briefly on one spatial task after ten minutes of a sonata. The press deleted "36," "briefly," "one spatial task," and "college," and by 1998 the governor of Georgia was budgeting classical CDs for every newborn in the state. Note the chain: no infants appeared anywhere in the study. **What the evidence says.** Meta-analyses (Chabris,

1999; Pietschnig et al., 2010) find a tiny-to-nil effect best explained by arousal and mood — anything you enjoy works as well as. There is, so to speak, an equally valid milkshake effect. **Verdict.** = Dead. Mozart is a composer, not a supplement. The enjoyment was already the payment.

• The 10,000-hour rule

The claim. Ten thousand hours of practice makes anyone an expert at anything. **Why it spread.** Ericsson, Krampe, and Tesch-Römer (1993) found elite violinists had accumulated *on average* about ten thousand hours of deliberate practice by twenty; Gladwell (2008) rounded a mean into a law with a number that fits on a poster. **What the evidence says.** Macnamara, Hambrick, and Oswald's 2014 meta-analysis: deliberate practice explains about 12% of performance variance overall — ~26% in games, ~21% in music, under 1% in professions. The average hid an enormous spread; hours are an entry fee, not a purchase; and Ericsson himself rejected the "rule" (an average is not a threshold, and *deliberate* practice — at the edge, with feedback — is not the same as hours). The research the myth mangled is alive and well; it is Chapter 15. **Verdict.** = Dead. The number was a mean. The magic was the feedback loop.

Six graves, one pattern: each myth took a true, narrow result and sold it wide. The autopsy skill is itself transferable — it is 6:35's move in a lab coat: check what the aggregate hides, ask what *n* was, what task, how long the effect lasted, and who is selling. The next beautiful claim is already on its way.

Further Reading

Seven doors, honestly labeled.

- **Martin Gardner, the "Mathematical Games" corpus (1956–1981).** The greatest amplifier in the genre's history; start with *The Colossal Book of Mathematics* (2001) and budget for the disappearance of several weekends.
- **George Pólya, *How to Solve It* (1945).** The founding document of heuristics-as-curriculum; Part III of this book is one of its many descendants.
- **Raymond Smullyan, *What Is the Name of This Book?* (1978).** The knights-and-knaves ladder that walks a patient reader, one tavern riddle at a time, all the way to Gödel.
- **H. E. Dudeney, *The Canterbury Puzzles* (1907) and *Amusements in Mathematics* (1917); Sam Loyd, *Cyclopedia of Puzzles* (1914).** The golden age at the source — Dudeney for elegance, Loyd for showmanship, both read best wearing the provenance goggles Chapter 2 issued.
- **Art of Problem Solving — artofproblemsolving.com.** Textbooks, forums, and the nursery of most modern olympiad medalists; the working ladder from school mathematics to Chapter 8–11 territory.
- **Project Euler — projecteuler.net.** Hundreds of problems engineered so brute arithmetic fails and only an insight runs in under a minute; mathematics and programming in a single gym.
- **The ARC Prize — arcprize.org.** Chollet's grids, built to be easy for humans and hard for machines. Do twenty yourself and feel, first-hand, which side of Part IV's ledger you sit on (standings there dated as of July 2026).

FURTHER READING

Colophon

This book was made the way it tells you to work.

The fifty-puzzle canon was compiled from primary sources, and the corrected attributions were stated in the text rather than footnoted away: the 15 Puzzle belongs to Noyes Chapman, not Sam Loyd; the zebra belongs to *Life International*, not Einstein; the Gauss anecdote is told as the legend it is. Sixteen of the canon's solutions were machine-verified by exhaustive enumeration or simulation in July 2026 — breadth-first search for the river crossings, all twenty-four cases of the twelve balls, the harmonic sum of the hundred prisoners, the uniqueness of 6210001000 — because a solution is not "verified" until something that cannot want it to be true has agreed. The worked solutions of Part III were re-derived end-to-end during compilation, and where two independent derivations existed, both were run and made to agree. Every claim about artificial intelligence carries its date — July 2026 — on purpose: those sentences will age, and the date is the difference between a snapshot and a pretense.

The book was composed by the AICE team, July 2026 — humans and machines in the division of labor the book itself recommends: the machines enumerated, checked, and never got tired; the humans chose the puzzles, the order, and what mattered.

One box has stayed open since the first page.

At the gate of Chapter 1, the Sphinx asked: *what goes on four legs in the morning, two legs at noon, and three legs in the evening?* Sixteen chapters and three appendices, and no answer — a spoiler discipline even we found hard to keep.

The answer is a human being. Crawling on four limbs in the morning of life, striding on two at its noon, leaning on a stick in its evening — one creature, three shapes, a single day standing for a lifetime. Oedipus solved it with the moves you now own: refuse the literal legs, re-bind the words, change the timescale. The Sphinx destroyed herself, as gatekeepers do when the gate is truly passed.

And notice what the answer was. Not a god, not a monster — the animal that changes how it moves and knows it. The oldest recorded riddle is about its own solver: the creature whose whole life is re-representation. Four thousand years later, her newest descendant sits in a benchmark showing colored grids to machines, still sorting minds at the gate. It is her only job. It has always been steady work.

The box is closed; the book is done; the gate, as always, is open.

Go on through.