Handicapping, Tested

Handicapping, Tested

A Lab Notebook for the Casual Horseplayer

This book was produced by an AI

A note before the first race

How to read the numbers

The one-page standard

Chapter 1 — The market is the best handicapper you know, and it still loses

Where the fifth of your money goes

Likely is not the same as worth it

What the crowd knows, and what it cannot do

The three lessons of the favorite

Chapter 2 — The morning line is a guess; the tote board is a vote

What the morning line actually is

What the tote board actually is

The morning-line order, stated as a rule

"Beat" means three things, and you must say which one

The first challenges: small counts, same answer

The large count: pace and separation rankings against the anchor

The wider record: committees, models, and thresholds

Why the guess holds up — and where it breaks

The three lessons of the two prices

Exercise: read the order first

Chapter 3 — Takeout: the tax that decides everything

One pool, worked in full

The rates, stated as what they are

Breakage: the second, quieter fee

Break-even: the one formula in this book

The crowd-matching machine, and the 25% arithmetic
Place and show: the same fee on a smaller prize
The same opinion in two pools: a decision, worked
What takeout does not do
The three lessons of the fee
Exercise: price your last twenty bets

Chapter 4 — Finding is not proving

Three ways a true-looking result can be false
What "out of sample" actually means
The model that knew too much: LOGIT_ALL
The family that reversed: FAIR thresholds
Why the lab freezes rules
What this means at your kitchen table

Chapter 5 — Better ranking, worse bets

Four claims, kept apart
The arithmetic that sits between order and money
The break-even line, in dollars
The worked example: our own record, taken apart
Calibration: the third claim, and the quietest failure
What to do with a ranking you trust
The closing discipline

Chapter 6 — Speed figures: what they measure and what they hide

What a figure actually is
The Sovereignty file
The control check: did the figure at least predict him?
Is the variant worth its trouble?
The rebuilds that failed
The constant underneath: an audit
How to read a figure on Saturday

Chapter 7 — Pace is a shape, not a number

Two numbers and a picture

The vocabulary of shape

The comfort zone, and a warning label

The sketch, tested as a sketch

The big test: do pace interactions reorder anything?

Two more ways to ask, two more answers

What pace is for

Chapter 8 — Class: real, and mostly already priced

What "class" means when nobody is selling you anything

Pars: the typical number for the race

How the lab built the pars

The diagnostic: class adjustment describes favorites better than anything else we tested

The pick rule: take the highest-CAP horse in the race

What a casual player should actually do with class

The three lessons of class

Chapter 9 — Consistency: the dependable horse and the improving horse

The measurements: spread and slope

The dependable horse: a shallow slope that vanishes

The improving horse: a true diagnostic and the program's worst rule

The peak that was not a fluke: the V3 screen

What a casual player should actually do with form

The three lessons of consistency and improvement

Chapter 10 — The cost of complexity

The specimen on the table

What the machine bought

Each factor is another ticket in the noise lottery

Each override re-opens a closed question
The whole becomes impossible to falsify
Four more parts that did not earn their place
The counter-discipline

Chapter 11 — The efficiency myth
First, separate the claims
The corpus test: the ratio in its first form
The diagnostic that runs backwards
The caveat this chapter owes you
The exact form: same ratio, our own cards
One more rescue was tried: efficiency, gated by pressure
What to do with a ratio at the track

Chapter 12 — Lone speed and other stories that invert
Write the story down as a rule
What the count showed
Why you cannot flip a failed screen
Why the story felt so true
The second story: the closer who leaves too much
What inconsistent costs you
How to test a story yourself
The point of the two stories

Chapter 13 — One-factor wonders
The counting rule, stated exactly
The population
A large share of every field
What a wonder wins, and what it pays
Which factor, and how many hits
The checks: a second definition and a second population
Why the best single number feels like more than it is
What to do with a one-factor horse

Chapter 14 — Pressure and separation: two measurements that tell the truth

The first measurement: who else wants the front

The second measurement: how far ahead the best figure stands

What these measurements claim, exactly

The rules the laboratory built on them — and what happened

What would promote a description to a prediction

How a casual player uses a description

The three lessons of pressure and separation

Chapter 15 — Committees of numbers

The three scoreboards

The laboratory's own committee, on all three

The visiting committee

The six disagreements, and the one useful thing

The scoreboard that was never filled in

The three lessons of committees

Chapter 16 — Track bias without superstition

The worked count: Keeneland at the close of its fall meet

What fourteen days showed

The shorter window, and the single day

The meet as a whole is not a bias

How the lab reads bias every day

Four claims, and which one a bias line makes

Reading a bias line like a casual player

Chapter 17 — The races to leave alone

No data is not neutral data: the Unknown in the field

The turf cells the lab will not pick

The price ceiling: no leg over 8-to-1

A warning sign that is not a rule: the horse whose form is on the wrong surface

Three worked passes
What passing costs and saves across a season

Chapter 18 — Money: units, stops, and records
What a staking plan can and cannot do
The unit: bet a fixed, small piece
The daily cap: the day has a budget, not just the bet
What a season costs: read this before the first bet
Resizing: once a week, in one direction when it matters
The drawdown ladder: decide the stops while you are calm
Records: the section that makes the rest of the book possible
The rules on one page

Chapter 19 — Test it yourself
Part 1: Write the rule in advance, in full
Part 2: Count every race it touches
Part 3: Compare against just betting the favorite
Part 4: Keep the result file, whatever it says
Part 5: Let it die if it loses
The worked example: does the most efficient closer beat the favorite?
Three ways this test could have been cheated — and what the freeze did about each
Your turn: the blank worksheet
The standard, one last time

Appendices
Appendix A: The closed list
Appendix B: Glossary for the casual player
Appendix C: Where to get past performances and charts

Handicapping, Tested

A Lab Notebook for the Casual Horseplayer

ThoroughWiz

This book was produced by an AI

This book was produced by an AI handicapping lab. That statement belongs on the title page, not in the fine print, so it stands here at the front: the tests in this book were run by an artificial intelligence operation, ThoroughWiz, working in a strict, deterministic environment set up as a handicapping laboratory, and the words were written from its records. The lab's work has been independently audited by other large language models. No individual model's name appears anywhere in this book, because no individual model is the author of the record. The record is.

What the lab does is narrow. It takes ideas about horse racing — old ones from books, new ones from its own work, ones that sound obvious and ones that sound clever — and it tests them against records of real races before it prints a word about them. Every major idea in this book went through that process. The ones that passed are here. The ones that failed are here too, with their numbers, because a lab notebook that only prints its successes is not a notebook. It is an advertisement.

Here is the promise this book makes, in the exact words the lab adopted before the drafting began:

> *"Every major handicapping idea in this book faces the same test: actual race results, a defined rule, and a measurable comparison. Some ideas survived. Others failed. The failures are published alongside the successes. No secret system. No guaranteed winners. No manufactured confidence. Just evidence showing what handicapping methods actually accomplish, what they don't, and how to tell the difference."*

The rest of this front matter explains how to read what follows: the numbers, the labels, and the standard every idea had to meet.

A note before the first race

Let us be clear about what this book is not.

It is not a picks service. Nothing in these pages tells you which horse to bet in a race that has not been run yet, and nothing in them is a tip. Where the book describes the lab's own daily selections, it describes a tested method and its full record, including its losses. It does not hand you tomorrow's card.

It is not a system for sale. There is no second book, course, or subscription that completes the method. The method is in this book, and most of it fits in one sentence: write an idea down, count what it did on every race it touched, and keep it only if the count says to.

It is not a promise of profit. You will find losing numbers in this book, including ours. Our own selections, on the only count we are permitted to publish — our own settled product record — lost money, and lost more of it than the favorite lost on the same races. Chapter 1 shows you that count in full: on the filed cards of November 2–4, 2026, 95 races carried a filed selection; 90 of them were bets, after scratch refunds; 25 won; $180 staked at two dollars a race returned $101.20, a return of −43.78%. Three days of racing prove nothing in either direction, and the book says so wherever it prints these numbers. What they prove is that we publish our record as it stands, not as we would like it. If a book about betting horses will not show you its own losing numbers, put it down.

What this book offers instead is a set of tested answers to questions casual players ask every day. Does this kind of horse really win more often? Does that famous angle pay for itself? Where does the money actually go? Some answers will confirm what you already suspected. Some will reverse it. Each answer comes with the size of the test behind it, so you can judge how much weight it can carry.

You do not need a computer to use any of it. Every calculation in this book can be redone with a pencil, and the final chapter shows you how to run the same kind of test yourself, on your own track, with free results.

How to read the numbers

Every test in this book reports the same three things. Learn to look for them, here and everywhere else you read about betting, because a claim without them cannot be checked.

The race count. This is the number of races the test covered. It tells you how much evidence stands behind the result. Eight races is a story. Eight hundred races is a pattern worth a second look. Eight thousand races is a fact you have to work around. When a result in this book rests on a small count, the text says so, and you should hold it loosely.

The win rate. This is the share of bets that won, stated as a percentage. If a method picked the winner in 31 of 100 races, its win rate is 31%. A win rate by itself tells you nothing about money. A method can win often and lose, by collecting small payoffs on short-priced horses. It can win rarely and come out even, if the payoffs are large enough. The win rate answers one question only: how often was it right?

The return on $2 win bets. This is the money line, and it is the one that matters. Every test in this book stakes the same imaginary bet: two dollars to win, on every race the method touches, no exceptions and no skipping the races it got wrong. At the end we count what came back. If $2,000 went in and $1,600 came back, the return — the ROI, for return on investment — is −20%. The method lost twenty cents of every dollar. If $2,100 came back, the ROI is +5%.

Two dollars is a deliberate choice. It is close to the smallest bet most tracks accept, it is the base our lab keeps its own books at, and it makes the arithmetic easy to check. Double it and every number in this book doubles. The percentages do not change.

Two verdicts appear throughout the book, and they mean exactly what they say.

Supported means the idea was written down in advance, tested on every race it applied to, and beat the standard it had to beat — usually the favorite, for reasons the one-page standard below explains. Supported does not mean profitable. An idea can beat the favorite's return and still lose money, and in this book some do. Supported means the evidence said the idea was real, at the size of test behind it, under the gates set before the test ran.

Not supported means the idea failed that process. It may have lost to the favorite. It may have won too rarely, or at prices too short, to cover its losses. It may have worked on a small count and stopped working when the count grew. Not supported is not an insult to the idea or to whoever proposed it. Some of the ideas in this book that came back not supported were ours, and a few of them we liked a great deal. The count does not care what we liked.

One more habit to build now, because it will save you confusion later: when this book gives you a percentage, ask "of what, over how many?" A win rate of 40% over 50 races and a win rate of 40% over 5,000 races are the same number and very different evidence. The first can happen by luck in a single good month. The second cannot.

The four evidence classes

Not every number in this book stands on the same ground. A result measured on the lab's own collected records is a different kind of evidence from a result measured on a research dataset the lab was given access to, and both differ from a piece of racing knowledge nobody in the lab measured at all. So every chapter carries a label, and the label travels with the number wherever the book cites it. There are four classes.

OWN means the result was rebuilt on ThoroughWiz's own collected records: the lab's store of 119,105 past-performance lines, and the product cards it published and settled day by day. OWN states where the numbers came from. It does not state that they are cleared for publication: publication rights for content derived from those records are to be verified before print, and the book treats that verification as a separate step from the measurement.

CORPUS means the result was measured on a research corpus: a large set of historical race records the lab was given access to for research. The corpus is a research dataset, and its statistics are publishable only with the dataset owner's publication terms. The lab does not hold those terms, so in this edition every corpus statistic has been removed: the tests are described, their rules and verdicts stand, and their numbers do not print. Where

a chapter's tests were measured on the corpus, the chapter says so — and a result rebuilt on the lab's own records changes the evidence class to OWN, as the book's own selection record was rebuilt for Chapter 1.

GENERAL means established racing knowledge, with no lab result attached. The takeout ranges in this front matter are GENERAL. So is most of what any experienced player could tell you about how a tote board works. The book labels these passages so that you never mistake a piece of shared background for a measurement.

INDEPENDENTLY REPLICATED is the strictest class, and the book claims it only where it has happened. It means a dataset independent of the original research reproduces the finding. Rebuilding a result from our own stored records is not independent replication — it is the same evidence, re-counted — and no chapter may claim this class on that basis. If a finding in this book carries the label, it earned it from outside data. If it does not carry the label, it has not.

Read the labels as part of the number. A −18% return labeled CORPUS and a −18% return labeled OWN are the same arithmetic on different ground, and the ground is part of the claim.

Four different claims

Handicapping talk fails most often at one quiet step: it slides from one kind of claim to another without noticing. This book keeps four claims separate, grades them separately, and never lets one stand in for another.

An **association** claim says two things move together. Favorites whose recent figures were all earned on a different surface win less often than other favorites. That is an association. It describes a population. It does not, by itself, name a bet.

A **ranking** claim says a method orders a field better than some other ordering — that its first choice wins more often, or that the actual winner sits higher in its list. Ranking is about order, not about price.

A **probability** claim says a method's stated chances match reality: that the horses it calls 30% shots win about 30% of the time. A method can rank a field well and price it badly, or price it honestly and rank it no better than the crowd.

A **returns** claim says the money works: that betting the method, at the prices actually offered, returns more than the stake after the takeout has taken its share — or, in the weaker form this book uses most often, that it loses measurably less than the benchmark loses on the same races.

Each step up that ladder is harder than the one before, and evidence for a lower claim is not evidence for a higher one. A measurement can describe a favorite's weakness precisely and still fail to name the horse that beats it — that pattern recurs so often in the lab's record that it has its own chapters. Whenever this book says an idea worked, it says which claim it means. Hold every other book, tip sheet, and system seller to the same discipline, and most of their claims will quietly shrink to the rung they actually earned.

A companion distinction runs through every chapter: **beating a baseline is not the same claim as making money.** On the lab's publishable record, our own selections did not even beat the

baseline: they lost 43.78% while the favorite lost 38.88% on the same races. Three days is far too small a count to grade a method by — and that, too, is part of the finding.

Backtest and prospective

Every result in this book carries one more label, stated in words: was the rule tested on races that had already been run when the rule was written, or on races run after the rule was frozen?

A **backtest** runs a rule over historical races. It is the necessary first test — an idea that loses on races you can count completely will not improve on races you cannot — and it is exploratory evidence. A rule can be shaped, knowingly or not, by the very races that judge it: its designer saw those results, chose its cutoffs with them in view, and discarded the versions that looked worse. A backtest result therefore says "this rule is worth watching," not "this rule has performed."

A **prospective** result comes from a rule frozen in a numbered, uneditable form before the races it is judged on were run, then left alone while new races accumulate. Prospective evidence is slower and smaller, and it is the only kind that measures performance rather than fit. The lab's own selection method carries both labels in Chapter 1, stated plainly: its headline numbers are a backtest, from a method refined across earlier versions with knowledge of historical races; its prospective record — published cards, settled in public, with standing watches judged against tripwires written in advance — began on November 2, 2026, and is days old at this writing. The book will not pretend otherwise, and it grades every other method by the same split.

The takeout, in one page

One number sits underneath every test in this book, and you need it before any of the tests make sense.

Horse racing is not a bet against the track. When you bet a race, your money goes into a pool with everyone else's money on that race. The winners split the pool. But first, the track removes a fixed share of the pool to pay for the races, the purses, and the operation. That share is called the **takeout**.

In a typical North American win pool, the takeout is roughly 15 to 17 cents of every dollar wagered. The exact rate varies by track, by state, and by type of bet, and Chapter 3 works through the differences in full. For now, hold the simple version: bet $100 across a win pool, as part of the crowd, and about $83 to $85 is what the winners — including you, when you win — divide up. The rest left the pool before the race was run.

A second, smaller charge hides inside the payoffs. Payoffs are rounded down to the nearest dime or nickel, and the track keeps the rounding. This is called **breakage**. It is small on any one bet. It is steady, and it only moves in one direction.

Takeout explains the central fact of this book, which Chapter 1 puts on the table: the favorite — the horse the crowd likes most, backed by the pooled judgment of every bettor in the pool — wins about one race in three and still loses money on it. On the lab's own cards of November 2–4, 2026, the morning-line favorite started in 101 races, won 27 of them (26.7%), and returned −45.00% — a hard three days for favorites, and a reminder of how violently a small count swings. The crowd is good at picking winners. The crowd is not good enough to pay a 15-to-17-cent fee on every dollar and come out whole. Nobody

is, most of the time, and Chapter 3 shows the arithmetic of why: to break even against that fee, you do not need to match the crowd. You need to beat it by a wide and specific margin.

Every method in this book is graded against that reality. A method that loses 18 cents per dollar has not made you money. It has lost two or three cents less per dollar than the crowd loses, and at honest counts, that is what an edge looks like. Keep that scale in mind and the numbers in this book will read correctly — as measurements, not as promises.

The one-page standard

Everything in this book passed through one filter. It fits on this page.

An idea earns its place here only if it worked on past races, counted in full, before anyone relied on it going forward.

Each part of that sentence does work.

Worked means it beat the benchmark named below, at the standard set before the test ran. An idea that merely sounds right, or that a respected name endorses, or that won last weekend, has not worked in this sense. It has been described.

On past races means the test runs on races already run, with their results known. That sounds backward — surely what matters is the future? The reason is discipline. An idea tested on the future can be adjusted as it goes, quietly, until it looks good. An idea frozen and then run over thousands of finished races cannot be adjusted at all. If it survives that, it has earned

the right to be tried forward. If it does not survive, no forward trial can rescue it, because a method that loses on races you can count completely will not improve on races you cannot.

Counted in full means every race the idea touched goes in the count: the losers, the short prices, the bad days, the races where the idea looked foolish. The most common way a betting idea deceives its owner is not fraud. It is selective memory — the wins retold, the losses forgotten, the awkward races quietly reclassified as races the system "didn't really like." A full count makes that impossible. This book's Appendix A lists the ideas our lab tested and discarded, with their verdicts, for the same reason: the failures are part of the record.

Before any test runs, the lab writes the idea down in a fixed form: the rule, stated so a stranger could apply it; the races it applies to; the benchmark it must beat; and the margin it must beat it by. This is called **pre-registration**, and in plain language it is just a promise made in writing before the results come in. It stops the quiet editing that turns a losing idea into a winning story — the shifted cutoff, the dropped track, the "that race doesn't count." If the rule changes after the results are known, it is a new idea, and it goes to the back of the line for its own test.

The benchmark is the **favorite**, and behind it the morning line's plain order. Every idea in this book had to beat simply following that order in the same races — in the lab's practice, the morning-line favorite and the selections built directly on the morning line, since the morning line is the pre-race form of the crowd's judgment that a player actually holds in hand. Chapter 1 shows why this benchmark is the right one. The favorite is the pooled judgment of everyone who bet, including professionals and computer teams wagering sums a casual player will never

match. It wins about one race in three. It still loses about a fifth of the money bet on it, because of the takeout described above. Beating that return is a low bar in one sense — the bar itself loses — and a demanding one in another, because it means beating the crowd's pricing, not just picking winners.

Five refinements to that standard were earned the hard way, each by a specific reversal in the lab's own record, and the book follows all five. They are stated here in plain language because later chapters lean on them.

Name the ground. Every verdict carries the label of the population it was measured on — which corpus, which tracks, which year — wherever the book cites it. A result measured on one population and quoted as if it were measured on another is a different, weaker claim wearing the first one's clothes. And if the same test, run on two populations the lab keeps, gives different gate verdicts, the question does not get averaged away. It goes back to a human decision, on the record.

Two grades, not one. An idea that passes can pass at two different strengths. At the stronger grade, its money interval — the range the true return plausibly sits in, given the size of the test — clears zero entirely: the economics are proven, not just pointed at. At the weaker grade, the pick-rule evidence is real (it names winners better than the benchmark, head to head) but the money claim is unproven, and the idea is adopted as an annotation: a way of reading a race, with no wagering meaning. Winning a head-to-head duel never, by itself, promotes an idea to the stronger grade. The lab's own selection method sits at the weaker grade, and Chapter 1 shows the interval that puts it there.

Check the concentration. Before an idea is graded, the lab asks where its edge lives. If most of the advantage comes from a single factor, a single kind of race, a single track, or a single band of prices, the edge is presumed to be a spike — a lump of luck in one corner of the data — until the concentration is explained. Real edges tend to be spread out. Spikes are what multiple testing manufactures for free, and Chapter 4 shows the machinery.

Account for what a rule displaces. Any rule that overrides another rule — a cap that replaces a pick, a gate that blocks one — changes two horses' fates, not one. When the lab adopts such a rule, it sets a standing watch on the displaced horses at the same moment, with a numeric tripwire written in advance: if the displaced horses outperform the horses installed over them by the stated margin over the stated count, the rule's exception comes back for a decision. An override without a watch is an untested claim hiding inside a tested one.

State what is new. Before a test runs, the proposal states what information the idea uses that the lab's existing measurements do not already contain. A new combination of genuinely new information is a new method. A re-weighting of the same factors the lab already uses is a variant of the existing method, and it is tested under the stricter rules for variants — it does not get to present itself as a discovery. Several chapters turn on this distinction, because it is the difference between learning something and rearranging something.

For the record, and because this book holds us to the same standard it holds everyone else: our own selections, on the record we are permitted to publish — the filed cards of November 2–4, 2026 — returned −43.78%, against the favorite's

–38.88% on the races where both can be compared. On that count, the gap runs 2.88 points against us. The count is three days long. It grades nothing, and this book does not ask it to. The larger counts behind the method's design were measured on a research dataset whose numbers we may not print; what we may print is the record we publish nightly, and it starts here, losing. That is the honest product of testing, counting, and discarding: a method graded at the annotation level described above, with its prospective record still young and published in the open as it grows.

That is the standard. Ideas that meet it are in this book as supported. Ideas that do not are in this book as not supported, with their numbers attached. Ideas that were never tested are not in this book at all.

Now to the races.

Chapter 1 — The market is the best handicapper you know, and it still loses

Pick a Saturday. Pick any track you like. Walk the card race by race and write down the favorite in each one — the horse the betting crowd sends off at the shortest price. Bet two dollars to win on every one of them, all day, no opinions, no handicapping, no hunches.

Here is what happens, and it happens with the regularity of a train schedule. Your horse wins a minority of its races. At the end of the day you count your money and find that a large part of it is gone.

That is not a guess about your luck. It is a count. Our laboratory keeps records on every idea it tests, and the favorite is the first line in the ledger because every other idea has to beat it. On the lab's own settled product record: the filed cards of November 2–4, 2026 — 111 races, the morning-line favorite started in 101 races and was scratched in 10. It won 27 of those 101 races, a 26.7% strike rate. A two-dollar win bet on every one of those favorites staked $202.00 and returned $111.10. The net return is −45.00%. That record is three days old and three days small. It proves nothing on its own about the long run. It sits in the same direction as the general racing knowledge this book states as approximate: over large counts the favorite wins about one race in three and blind favorite money loses about a fifth, after takeout. The crowd's chosen horse, backed blindly, loses money. The rebuilt record loses more, on a sample too small to settle the rate.

Sit with that for a moment, because it is the most important fact in this book. The favorite is not a fool's bet. The favorite is the pooled judgment of everyone who bet the race: the regulars who live at the track, the sharp players with their own figures, the computer teams that wager millions a year with models built by people who do this for a living. When the gates open, the horse the money liked best wins more often than any other single horse. No individual handicapper's first pick, ours included, does better at naming winners. The market is the best handicapper you know.

And the best handicapper you know loses money.

Where the fifth of your money goes

Part of the loss is mechanical, and it helps to see the machinery. Horse racing is not a bet against the track. It is a bet against the other players, with the track taking a fee for running the pool. That fee is called the takeout. In a typical North American win pool, the track removes roughly 15 to 17 cents of every dollar before anyone is paid. Breakage — the rounding down of payoffs to the nearest dime or nickel — takes a little more.

Do the arithmetic on a perfect handicapper. Suppose you could pick winners exactly as well as the crowd does, at exactly the crowd's prices. You would win a third of your bets, collect the same payoffs the favorites collect, and lose the takeout on the whole flow of money. Skill that merely matches the crowd is a machine for paying the fee. To break even in a win pool you do not need to be as good as everyone else combined. You need to

be better than the crowd by roughly the size of the takeout — on the order of a quarter again as accurate, in money terms, as the pooled judgment of every other bettor in the pool.

That is the real shape of the game, and almost everything written about handicapping becomes clearer once you see it. You are not trying to find the winner. A third of the time, the winner is standing in plain sight at 8-to-5. You are trying to find the races where the crowd's *price* is wrong — where a horse's true chance of winning is meaningfully better than the chance its odds imply. The winner and the good bet are different animals. The first question of handicapping is "who is most likely to win this race?" The crowd answers that question better than you do. The second question is "who is priced wrong?" That is the only question this book is about.

Likely is not the same as worth it

A short story makes the difference plain. A horse goes off at even money — a dollar of profit for every dollar bet. The crowd's price says it wins about half the time, after the takeout does its quiet work. Suppose the horse is genuinely good and genuinely the best in the field, and it wins 45% of races like this one. It is the most likely winner. It is the right answer to "who wins?" It is also a bad bet: at 45%, a fair price is a little better than 6-to-5, and the board is offering you even money. Bet it a hundred times and you lose about ten dollars on every hundred, cheering for the best horse in the race the whole way down.

Now take its rival at 6-to-1, a horse the crowd gives one chance in seven. Suppose you are right, for reasons you can state, that it is really a one-in-five horse. It will lose four times out of five,

and it will feel like a loser most afternoons. Bet it a hundred times and you come out ahead, because the price paid you for more risk than you actually took.

Neither horse cares what it paid. Horses do not know their odds. But the two bets are opposites: one buys the likeliest outcome at a price that guarantees a slow loss, the other buys an unlikely outcome at a price that pays for its unlikeliness and then some. Casual players lose money mostly in the first way — not by picking bad horses, but by paying bad prices for good ones.

What the crowd knows, and what it cannot do

Here is the crowd's record in full, because this book will return to it in every chapter. The crowd ranks horses well. Its ordering of a field — favorite, second choice, third choice, and down — tracks reality closely enough that no simple rule we have ever tested improves on it by naming different winners. Our lab spent a season testing the standard alternatives: rankings built from speed figures, from pace figures, from combinations of both, from the wisdom of expert panels and voting systems. Chapter by chapter, this book walks through those tests. The pattern never changed. The alternatives named winners less often than the morning line's order did, or named the same winners at worse prices.

Our own operation is not exempt from this lesson, and it would be dishonest to exempt it. We publish selections every day, made by a consensus method: it scores the morning line's top four on five factors, lets the votes compete, and applies a price

cap before it files a pick. Here is its record with the disclosure a claim like this owes you — every number, not the flattering ones.

The population is the lab's own settled product record: the filed cards of November 2–4, 2026 — 111 races. That is three days. The record is days old, the sample is small, and the uncertainty around every rate in this section is wide. Three days prove nothing either way. The record is published nightly as it grows, and this book prints the record as it stands at the rebuild.

On those cards, 95 races carried a filed selection. A scratched selection is a refund, not a bet, so the §26 selection book — a $2 win bet on the consensus selection in every race where one was filed — placed 90 bets. It won 25 of them, a 27.8% strike rate. It staked $180.00 and returned $101.20. The net return is –43.78%. It lost money.

The product also settles one official leg per race, with its filed backup chain when the pick scratches. That official-pick book carried 95 legs. It won 27 of them, a 28.4% strike rate. It staked $190.00 and returned $118.20. The net return is –37.79%.

The headline book as published nightly adds the §16 counter-move legs to the official legs. It carried 108 legs. It won 28 of them. It staked $216.00 and returned $133.40. The net return is –38.24%.

The consensus step changed the official pick in 12 of the 95 races. The price cap then capped 5 of those changed picks back to the favorite.

The comparison that matters is against the favorite on the same races. In the 85 races where both the selection and the morning-line favorite started, the selection won 24 races, a 28.2% strike rate, and returned $99.00 on $170 staked, a return of −41.76%. The favorite won 25 races, a 29.4% strike rate, and returned $103.90 on $170 staked, a return of −38.88%. The paired difference is −2.88 points — on this small a record the selection did not beat the favorite. Say so plainly: it did not. Read that at the right scale. Three days of filed cards cannot settle whether a method loses a little less or a little more than the chalk. The honest sentence is that three days prove nothing either way, and the record is published nightly as it grows.

How sure are we of any edge the method claimed in its research tests? Less sure than decimal places make it look. The lab ran a separate economics test on an earlier, four-factor form of the same consensus idea. Its measured edge carried a confidence interval that crossed zero. An interval like that includes zero. It includes the possibility that the true edge is negative. The edge was also spread unevenly: it sat in one price band, and the second half of the year accounted for more than the whole net, because the first half ran negative. The lab's ruling on its own method is the honest label, and this book adopts it as a category: ANNOTATION ONLY — a pick rule with evidence behind it and no proven wagering meaning.

One more disclosure, the one that matters most. This method was not written down in advance and then tested on untouched races. It was developed and refined across several versions while its builders knew historical races and their outcomes. Its research results are backtest results — exploratory evidence, a reason to keep watching, not a record of live performance. The

genuinely prospective evidence is younger and smaller: the method's rules are now frozen as a numbered version that cannot be quietly edited; our daily cards have been published and settled in public since November 2, 2026; and standing watches judge parts of the method on each new day's races against tripwires written in advance. That record is days old. It will take a season to mean anything, and this book will not pretend otherwise.

Hold on to the distinction this record forces, because the whole book runs on it: **beating a baseline is a different claim from making money.** A method can beat a baseline in a research test and still lose money in the filed book. An idea can also describe a race accurately, rank its horses better, or estimate their chances better — four different claims, and every chapter that follows grades them separately. Whenever this book says an idea worked, it will say which claim it means.

Anyone who tells you their method turned the crowd's loss into a large profit is selling you something. The numbers do not bend that far. They never have, in any count anyone has published with the races all included.

The three lessons of the favorite

This chapter asks you to keep three things, and the rest of the book is footnotes to them.

First: respect the ranking, question the price. The crowd's order is information — the best free information in the game. Start every race by asking what would have to be true for the price to be wrong, not by asking which horse you like.

Second: never confuse the likeliest winner with a bet. A bet needs a price that pays for the risk. The best horse at a bad price is the most expensive habit in racing.

Third: your opponent is the takeout, not the other players' horses. You cannot beat the fee by matching the crowd. You beat it — to the small extent it can be beaten — by betting less often, in the places where the crowd's pricing is measurably sloppy, and by refusing the bets where it is sharp. The chapters ahead map those places. Some of the map will surprise you. A few of the most famous roads in handicapping turn out, on a full count of races, to lead nowhere at all — and two or three plain measurements that nobody writes books about turn out to describe a race more honestly than the famous ones.

That mapping is the work. It begins with the sheet of numbers every player already holds: the morning line, and what it actually is.

Chapter 2 — The morning line is a guess; the tote board is a vote

Open any racing program to any race and you find two prices for every horse. One is printed in ink, hours before the race, next to the horse's name. The other does not exist yet when you buy the program. It forms in the minutes before the gates open, it moves while you watch, and it freezes only when betting closes. Players talk about these two prices as if they were the same thing at different times — an early price and a late price. They are not. They are answers to two different questions, made by two different processes, and mixing them up is the first pricing error a casual player can learn to stop making.

The printed price is the **morning line**. One person makes it: the track's handicapper, working the night before or early that morning, alone with the entries. The tote price is the **closing market** — the final odds when betting stops. Thousands of people make it, with their money, including professionals and computer teams whose wagers arrive in the final seconds. The morning line is a guess. The tote board is a vote. This chapter takes each one on its own terms, defines exactly what it would mean for anything to beat either one, and then shows you the test record behind a plain finding that surprises most players the first time they hear it: in our lab's work, the simple order of that one person's morning guess — first choice, second choice, third choice, straight down the page — kept its job as the starting ranking against every ranking built from speed figures, pace figures, and combinations of the two that we tested. The guess, read as an order and nothing more, beat the figures. Read that sentence at the right scale. It is a claim about ranking. It is

not a claim about making money, and by the end of the chapter you will be able to say exactly why those are different claims, with the denominators attached.

What the morning line actually is

Start with the job description, because the job explains the product. The track handicapper who sets the morning line is not asked, "Which horse will win?" The assignment is narrower and stranger: "Tell us, before a dollar is bet, how you think the public will bet this race." The morning line is a forecast of the crowd, made by one person, on a deadline, the night before the crowd exists.

That assignment shapes every line you will ever read. The linemaker works from the same past performances you hold, plus local knowledge: which trainers the local crowd trusts, which jockeys draw money on their names alone, which kinds of horses this public habitually overbets and underbets. A famous trainer's first-time starter may get a shorter morning line than its form earns, because the linemaker expects the crowd to bet the name. A sharp-looking horse from a barn the local public ignores may get a longer line than its form earns, for the same reason in reverse. The line is not a pure reading of the horses. It is a reading of the horses filtered through a reading of the bettors.

The linemaker also works under a habit you can check yourself. Add up the chances the line implies across a field — a 2-to-1 shot implies roughly one chance in three, a 4-to-1 shot one in five, and so on — and the total usually lands a little above 100%, because a forecast of a pari-mutuel pool leaves room for the

takeout the pool will remove. The line is built to resemble a finished betting market, not to list each horse's true chance in a world without a fee.

Two limits follow. First, the linemaker writes before scratches, weather changes, and a final surface — a line written at 8 a.m. cannot know the 3-to-1 second choice will scratch at noon. Second, the linemaker is one reader. A good one is skilled, and the record later in this chapter shows the skill is real. But one reader misses things a market catches, and a market misses things one careful reader catches. Keep the two prices in separate boxes in your head.

What the morning line is not: it is not the track's tip — the track's fee comes off the pool either way, as Chapter 3 shows in dollars. It is not a promise about the closing price. And it is not an early, stale version of the tote odds, to be thrown away when real money moves. It is a separate measurement, made by a different instrument. Instruments get compared by testing them, not by assuming the newer reading replaces the older one.

What the tote board actually is

The closing market works by pooling, and the pooling is literal. Every win bet on a race goes into one pool. When betting closes, the track removes the takeout, and the remainder is divided among the winning tickets in proportion to how much each ticket staked. The odds on the board are not set by anyone. They fall out of the money: a horse with a large share of the pool shows short odds, a horse with a small share shows long odds, and every new bet nudges every price in the race.

Read the board with that machinery in mind and familiar sights change meaning. Odds that drift upward late do not mean the horse got worse. They mean money arrived for other horses faster. Odds that crash late do not always mean insiders know something; late money includes large computer-assisted wagers placed near the close, because those teams wait for the fullest picture of the pool, and one large wager moves a small pool a long way. A horse bet down from its morning line has been voted for. A horse ignored past it has been voted against. The vote is information. It is not a verdict from anyone who knows the result.

Chapter 1 gave you the favorite's record as a handicapper on the lab's own settled product record: the filed cards of November 2–4, 2026 — 111 races, a small sample, three days, days old, with wide uncertainty. On those cards the morning-line favorite started in 101 races, won 27 of them, and returned $111.10 on $202.00 staked, a return of –45.00%. Hold both halves of that record at once. The pooled vote ranks horses well; no casual player's opinion, and very little else, ranks them better. The pooled vote also prices its own top choice at a level that loses money. Ranking and pricing are separate skills, and the market is good at one of them.

One more separation matters before the tests. The morning-line favorite and the closing favorite are often the same horse. Often is not always: scratches remove horses, weather moves races off the turf, late money promotes a horse the linemaker ranked third. When this book reports a test of "the morning-line order," it means the printed order, applied by fixed rules to the horses who actually started — not the closing order, not a blend. Every

test in this chapter names which favorite and which races, because a result you cannot assign to a specific price, on specific races, is a story, not a result.

The morning-line order, stated as a rule

"The morning line works" is too loose to test. Our lab's baseline, called ML-BASE v1.0 in its records, is the morning line written as a rule a stranger could apply without judgment calls. The full specification sits in the lab's files, and its working parts fit in one paragraph.

Take the horses who actually start — post-time starters, scratches removed. Also-eligibles count only if they draw in. Rank starters by printed morning-line odds, shortest first. Break ties by program number, lowest first, so a tie never needs an opinion. List coupled entries — two horses bet as one — under each program number separately, and note the coupling. A horse with no morning line goes last, by program number. The output is just an order: first, second, third, on down. Pace and speed figures may be attached as labels. Under the rule, they never reorder it.

Notice what the rule does not do. It does not tell you to bet the first horse. It assigns no horse a probability. It does not say the second horse is a better or worse bet than the first. It ranks. That modesty is deliberate, and it is why the rule survived: every challenger that tried to do more had to prove the extra work earned its place against an opponent that does almost nothing and does it consistently.

In the lab's current protocol the same order still holds the anchor job: filings build on top of it rather than replacing it. A probability model prices races for specific filing decisions, but it never ranks and never reorders the anchor. A later consensus step scores the morning line's top four on several factors and can change which horse is filed as the selection — Chapters 5 and 10 take that method apart in full — but the pool it scores is the morning line's top four, the morning line is one of its factors, and its status, stated in Chapter 1, is a pick rule with evidence and no proven wagering meaning. On the rebuilt product record the consensus step changed the official pick in 12 of the 95 races that carried a filed selection, and the price cap capped 5 of those changed picks back to the favorite. The anchor underneath never changed hands. The rest of this chapter shows you the attempts to take it.

"Beat" means three things, and you must say which one

Handicapping talk uses "beat" the way weather talk uses "nice" — as if everyone meant the same thing. Before any test result in this book can mean anything, the word needs a definition with edges. Here is the definition this book uses, in full, every time. A method **beats** a baseline only on **identical races** — the same set of races, counted for both sides, with the **denominator stated** — and only in one or more of three named senses:

1. **Higher win rate.** On those identical races, the method's pick won a larger share of the time than the baseline's pick. This is a ranking claim: whose first choice is right more often.

2. **Better return on investment.** On those identical races, a $2 win bet on every pick returned more — or lost less — for the method than for the baseline. This is a wagering claim: whose picks paid better at the prices actually offered.
3. **Better paired result.** Race by race, compare the two returns directly: in each race, subtract the baseline pick's return from the method pick's return, then average those differences over the stated races. A second paired measure, the **duel**, counts only the races where the two picks differ, and nets the method's wins in those races against the baseline's wins. Paired measures are stricter than comparing two separate averages, because they charge every disagreement to the race where it happened instead of letting good months and bad months hide inside totals.

Three senses, three different claims, and a method can pass one and fail the others. A ranking can name more winners at prices so short that its return is worse. A method can win less often and return more, because its winners paid better. A method can post a better overall return while losing the duel — winning its money in a few races and losing the head-to-head count across hundreds of others. Whenever a later chapter says a method beat something, or failed to, the sentence will name the sense, the races, and the denominator. Hold anything you read elsewhere — and anything anyone tries to sell you — to the same standard. "It beats the favorite" with no sense named and no races counted is not a claim. It is a noise.

Two more disciplines ride with the definition. First, **association is not any of the three**. A measurement can track winners honestly without any rule built on it beating the baseline in any sense. Describing a race and picking its winner are different jobs; Chapters 11 and 14 return to measurements that do the first well and the second not at all. Second, **a tie does not promote**. A challenger that matches the order has added work and added nothing. The lab's promotion bar, written into the baseline's specification, says it plainly: a method that ties the baseline does not get promoted, because extra complexity has to earn its place. Keep that sentence.

The first challenges: small counts, same answer

The baseline earned its job in a round of independent checks on full past-performance fields, recorded in its specification. The checks were small — dozens to a few hundred races each — and this book reports them at that weight: as the reason the lab adopted the morning-line order as the thing to beat, not as proof that nothing can beat it. Small counts can point. They cannot settle.

The challengers were the standard figure-based ideas of the time. A blend that averaged a published power rating, a speed rating, and the morning line returned a lower win rate for its top pick than the order did on the same cards, with a difference close to the line where it could still be luck, and on the wrong side of it for the challenger. A switch rule that moved the bet from the favorite to the second choice won less often than always taking the favorite. A top-three list built from the published power rating caught the winner in its top three less often than the morning line's top three did, in two checks. A

pace-augmented version of the line itself caught the winner in its top three less often than the plain order did, and where the two orders disagreed on the winner, the plain order was right more often than the adjusted order. One notation in the specification needs flagging, not smoothing: the specification records several test sizes with a label its file does not define, so this chapter reports the pattern and does not dress those labels up as race counts they may not be. The pattern does not depend on that label. In every check, in the win-rate sense, on the same cards: plain order first, challenger second.

Read those checks the way Chapter 4 teaches you to read small samples. Any one could bend on a few hundred more races. Together they justified a working decision, not a law: start from the morning-line order, make every new idea challenge it on identical races, and require the challenger to win by enough that luck is a poor explanation. The large counts came later. They carry the weight.

The large count: pace and separation rankings against the anchor

The lab's six-method program, run in 2026 on its large 2023 research corpus, put fresh figure-based rankings against the morning-line order at a size the early checks never approached. The corpus numbers from that program were struck in the 2026-11-05 rebuild and do not print in this book. The verdicts and the qualitative findings stand, and they are reported here at that level.

Against the morning-line favorite reference, four pre-registered ranking rules were scored. Each rule reordered the field by a figure construction and took its own top horse as its pick:

- A **pace-efficiency ranking**, which ordered horses by how much of their early pace figure they retained late, won far less often than the baseline and returned worse. Its paired difference against the baseline was negative, with a confidence interval that included zero and worse. In the races where its pick differed from the morning-line favorite, it lost the duel by a wide margin.
- A **pressure-exposure ranking**, which looked for horses whose rivals lacked matching early speed, also won far less often than the baseline and returned worse. Its paired difference was negative, its interval included zero, and it lost the duel by a wide margin.
- The **combination** of the two — the program's priority experiment, efficiency gated by pressure — won far less often than the baseline and returned worse. Its paired difference was negative, its interval included zero, and it lost the duel by a wide margin.
- A **separation ranking**, which ordered horses by how far their figures stood clear of the field, won less often than the baseline but returned slightly better than the baseline on its point estimate, with a small positive paired estimate. It still failed, on the other two senses at once: its interval included zero, and its duel was heavily negative. Where it disagreed with the favorite, it lost the head-to-head count by a wide margin. Its smaller loss came from the prices of the winners it did catch, not from out-picking the favorite.

The lab's verdict on each of the four was the same: **NOT SUPPORTED**. The control count on the wider research population agreed. Every verdict kept its sign when the population widened. No result in this section flips when the population widens. That agreement matters, and this book will flag any later result where the two populations disagree, because a verdict that changes when the track list changes is a verdict about the track list.

Pause on the win rates, because they teach the chapter's central fact better than the returns do. The figure rankings did not lose narrowly on picking winners. They won at roughly half to two-thirds the favorite's rate. These were not careless constructions; each encoded a real handicapping idea, each was written down before the count, and the measurements inside them describe races honestly — later chapters show pressure and separation tracking the favorite's strength in a clean, graded way. As descriptions, the measurements work. As rankings — as answers to "name a different first horse" — they named horses who won far less often than the plain order's first horse. The morning line had already absorbed most of what the figures know. The crowd's one-person forecast of itself, made the night before, carried the speed, the pace, the class, and the connections in a blended form no single-figure ordering in the test could match — because the linemaker, forecasting the crowd, was in effect forecasting a market that itself reads all of those things.

The wider record: committees, models, and thresholds

The six-method program was one season's attempt. The lab's full battery on its research corpus records the other families of ranking ideas against the same standards, and a chapter that only reported the attempts that failed in one particular way would be cherry-picking. Here is the rest of the record, each item with its claim graded separately. Corpus numbers were struck in the rebuild; verdicts and qualitative findings stand.

The consensus book. The lab's own selection method — disclosed in full in Chapter 1 — is printed in this book only on the rebuilt product record. Population label: the lab's own settled product record: the filed cards of November 2–4, 2026 — 111 races. Small sample — three days, days old, wide uncertainty. On those cards, 95 races carried a filed selection; the §26 selection book placed 90 bets, won 25 of them, staked $180.00, returned $101.20, and returned −43.78%. The official-pick book carried 95 legs, won 27 of them, staked $190.00, returned $118.20, and returned −37.79%. The headline book as published nightly carried 108 legs, won 28 of them, staked $216.00, returned $133.40, and returned −38.24%. Against the morning-line favorite on the 85 races where both the selection and the favorite started, the selection won 24 races and returned −41.76% ($99.00 back on $170), and the favorite won 25 races and returned −38.88% ($103.90 back on $170), a paired difference of −2.88 points. On this small a record the selection did not beat the favorite. Say so plainly. Three days prove nothing either way, and the record is published nightly as it grows. In the research tests the method's economics test, run on an earlier four-factor form, measured a positive paired point estimate with an interval

that crossed zero, with the edge concentrated in one price band and one half of the year. The lab's ruling stands: ANNOTATION ONLY — a pick rule with evidence, no proven wagering meaning. It also never replaced the anchor. It selects from inside the morning line's top four — starting from the line and declining most places where deviation was expensive, not out-ranking the line.

The top-four pool. A within-four test on the research corpus measured how much ranking work can be done inside the morning line's first four names. The winner sat inside those four in a large majority of the races — the pool the morning line hands you already holds most winners. A re-ranking inside the pool improved the four-horse book's return by a positive margin that fell short of its pre-set gate. The margin failed its gate. Real, positive, and too small, at the standard written before the count, to promote.

A locked statistical model. A logistic model the lab locked — frozen, never refit — was re-checked on its research population. On its holdout races it held a positive paired point estimate against the favorite and a positive duel, and it lost money. It failed its gates' ranking-margin leg and the profit leg. Its status is a paper candidate under a prospective judge — a status, not an adoption.

Threshold systems. A family of rules that bet only horses passing a modeled fair-price threshold ran four arms on the research corpus. Every arm lost money and every arm failed the family's clean-pass bundle. The verdict was NOT A CLEAN PASS. Thresholds changed how often the rules bet. They did not change the sign.

Stand back from the battery and the shape is consistent across methods, populations, and seasons of the lab's work. Rankings built from figures alone lost to the plain morning-line order in the win-rate sense by wide margins, and usually in the return sense too. Constructions that started from the morning line and worked inside it — a consensus over its top four, a pool re-ranking, a locked model — produced small positive paired differences in research tests that shrank under uncertainty checks and never cleared the bar for a wagering claim. The anchor held in the strong sense against figure rankings and in the strict sense against everything: nothing earned the right to replace the morning-line order as the starting ranking, on identical races, at the standards written before the counts. That is the whole claim of this chapter, and it is a ranking claim. Nothing in it says bet the morning-line favorite. The morning-line favorite loses money — on the rebuilt product record in Chapter 1, steeply, on a sample too small to settle its long-run rate. The anchor is where analysis starts. It is not where betting starts. Chapter 3 explains the difference in dollars.

Why the guess holds up — and where it breaks

Three reasons, all plain, explain most of the record. First, **the linemaker forecasts forecasters**. The morning line predicts a market, and a market aggregates every kind of reading — figures, pace, trips, stable moves, local knowledge — weighted by money. A forecast of that aggregate inherits most of its information before a single bet is placed. A ranking built from one figure family starts with less in the room.

Second, **the figures and the line read the same horses**. Speed, pace, and class readings are built from the same past performances the linemaker reads. A challenger that re-sorts the field by one reading mostly re-states information the line already weighed, then throws away the rest. For it to win, its one reading must be both strong and underused by the linemaker. On full counts, test after test says it is neither.

Third, **the order is a weaker claim than a pick, and weaker claims are harder to beat**. The order does not say by how much its first horse beats its second, or what either should pay. It only sorts. A figure ranking that sorts slightly differently and slightly worse has nowhere to hide; every swap is a race where one order was right and the other wrong, and the duels above count those swaps without mercy.

Where does the guess break? Where its job description says it must. It breaks when the race changes after the line is written: a scratch, a surface switch, a weather turn. It breaks where one reader's model of the local crowd is thin — full fields of lightly raced horses give any forecaster little to hold. And it breaks as a price, even when it holds as an order: a horse whose morning price looks generous is usually a horse the linemaker expects the crowd to correct. Treat a gap between the printed price and your own reading as a question — what does the room know, or what changed since the line was written? — not as a gift. Sometimes the line is simply wrong, and those races are worth your money. Finding them takes the price discipline of the next chapter, not enthusiasm about the gap.

Use the two prices for their two jobs. The morning line is your starting ranking and your map of what the crowd is expected to do. The tote board is the final vote and the only price you can

collect. A bet needs both: a reason, from the morning map, to think the crowd's expectation is wrong in a specific way, and a closing price that still pays you for being right after the fee. The first reading without the second is an opinion. The second without the first is a guess about other people's guesses.

The three lessons of the two prices

First: keep the prices separate. The morning line is one handicapper's forecast of the crowd, made before scratches and weather. The tote board is the pooled vote, computer teams included, made with money. Each is evidence about a different thing.

Second: "beat" is a defined word. Higher win rate, better return, or a better paired result — on identical races, with the denominator stated. By that definition, the plain morning-line order kept the anchor job against every figure-based ranking the lab tested, from small early checks to large corpus counts. No challenger that re-sorted the field by speed or pace figures beat it in the win-rate sense, and the one that returned better on its point estimate still lost the duel and carried an interval through zero.

Third: an anchor is not a bet. The morning-line favorite loses money, in the lab's rebuilt product record and in Chapter 1's. The order tells you where to start reading a race. The price tells you whether the race is worth money. Confusing those two jobs is how good handicapping turns into steady losses.

Exercise: read the order first

Take the next card you handicap — one track, one day, before you open a single past performance. Go race by race and write down the morning-line order for each race: program numbers in line order, first through last, with the printed odds beside them. Write down, in one short sentence per race, what the order claims: who the crowd is expected to make favorite, who it is expected to respect, who it is expected to ignore.

Now handicap the races as you normally do. Each time your reading moves a horse by more than two places against the written order, stop and write the reason: the specific fact — a figure, a pace shape, a trainer pattern, a trip note — that the linemaker missed or weighed differently. "I like this horse" is not a reason. "Second start after a layoff, with a late figure near the top of the field, against a favorite whose speed came on a track that favored it" is a reason.

After the races, score the sheet twice. First, the plain order: how often was its first horse the winner, and what would $2 win bets on it in every race have returned? Second, your moves: where you promoted a horse over the order's first choice, how often did each side win, and what did each side return? That is the duel and the paired return from this chapter, run on your own card, denominator in front of you. One card proves nothing — the counts in this chapter ran large for a reason. Run the sheet for a month and you will know, in the defined senses, whether your reading beats the guess. Most players who run it honestly learn the lab's first lesson: their moves win a few races the order missed, lose more races the order had right, and pay for the education until the moves get rarer and the reasons get better. That is the training working.

Chapter 3 — Takeout: the tax that decides everything

Chapter 1 left a fifth of your money missing and named the thief in passing. This chapter stops and does the accounting in full, because takeout is not one fact among many in this book. It is the fact the rest of the book is measured against. Every test in the chapters ahead was graded on returns after takeout. Every method that failed, failed against a fee it could not cover. Every method that survived, survived by a margin you can only judge if you know the size of the fee it was fighting. So we go slowly here, in dollars, with the arithmetic on the page where you can check it. Nothing in this chapter requires you to trust the lab. It requires a pencil.

The short version fits in one sentence: in horse racing you do not bet against the track, you bet against the other players, and the track removes a fixed share of the pool before the winners divide what is left. That share is the takeout. It is collected on every race, in every pool, whether you win or lose, whether you handicapped well or badly, whether the favorite wins or a longshot does. Rain does not change it. Your opinion does not change it. It is the one number at the track that never has a bad day.

One pool, worked in full

Take a single win pool with round numbers. Suppose the public bets $10,000 to win on one race, spread across the field. Suppose the takeout on this pool is 16% — a stated rate for this example, inside the range this chapter will give you shortly.

The track removes 16% of $10,000 before anyone is paid. That is $1,600. The winners divide what remains: $10,000 – $1,600 = $8,400. No winner ever sees the $1,600. It does not come out of anyone's payoff in a visible line. It comes out of the pool itself, so every payoff in the race is smaller than it would have been in a fee-free pool, by the same proportion.

Watch it land on one horse. Suppose $2,000 of the $10,000 was bet on the winner — one fifth of the pool. Each dollar on the winner claims a share of the $8,400: $8,400 ÷ $2,000 = $4.20 back for every $1 bet, stake included. A $2 ticket on that horse returns $8.40. In a pool with no takeout, the same $2,000 of winning money would divide the full $10,000: $10,000 ÷ $2,000 = $5.00 per dollar, and the same $2 ticket would return $10.00. The fee cost every winning ticket $1.60 — and it cost the losing tickets nothing extra, because they were already worth zero. This is how takeout hides. Losers lose their stake either way. Winners collect less than the pool's true proportions say they earned. The fee is paid, in full, out of the payoffs of the people who were right.

Now hold the same $10,000 pool and change only the bet type. Put the money in an exacta pool — pick the first two finishers in exact order — at a stated takeout of 24%, again inside the range given below. The track removes 24% of $10,000, which is $2,400. The winners divide $7,600 instead of $8,400. The same opinion, expressed in the second pool, starts $800 further behind in a $10,000 pool — 8 cents less of every dollar available to be won back — before a single horse has run. Nothing about your handicapping changed. The pool changed. This is the

sense in which the same opinion can be a reasonable bet in one pool and a poor bet in another, and it is why this book names the pool whenever it grades a bet.

The rates, stated as what they are

Takeout is set by law and by track policy, and it differs by jurisdiction and by bet type. The ranges in this paragraph are general racing knowledge, stated as approximate, and the draft header of this chapter flags them as such. Check the rate at your own track before you rely on it; the exact number is printed in the track's rules and in state or provincial regulations.

In North American racing, win, place, and show pools — the straight pools, where your horse has to finish first, in the first two, or in the first three — commonly carry a takeout of about 15 to 18 cents per dollar. Many win pools sit at 15 to 17 cents; some jurisdictions take more. Exotic pools — exactas, trifectas, superfectas, daily doubles, and the multi-race pick bets — commonly carry about 20 to 25 cents per dollar, and some pools in some places take more than that. The pattern is steady even where the exact numbers differ: the harder the bet is to hit, the larger the share the pool removes before paying it.

Two consequences follow from the pattern alone, before any arithmetic. First, a player who moves from win betting to exotics because the payoffs look larger is usually moving into a more expensive pool to chase the larger payoffs. The payoffs are larger partly because the bet is harder and partly because a smaller share of a larger-priced pool still prints a big number. Second, comparisons across bet types that ignore the rate are meaningless. A method that returns −18% in a win pool and a

method that returns −18% in an exacta pool have not performed the same. The second one fought a fee roughly half again as large and finished level. This book's own records keep the pools separate for exactly this reason — Chapter 2's win-bet numbers are win-pool numbers, and the exotic paper books in the lab's history are graded against exotic baselines, never against the favorite's win return.

Breakage: the second, quieter fee

Takeout is not the only removal. After the pool is divided, the payoff per ticket rarely comes out to an exact cent amount the track will pay. Payoffs are rounded down — commonly to the nearest dime on a $2 payoff, with the exact rounding unit varying by jurisdiction — and the track keeps the rounding. The kept fraction is called breakage.

Work it once and you will see it forever. Suppose the pool arithmetic says a winning $2 ticket should return $4.19. With dime breakage, the track pays $4.10 and keeps 9 cents. That is 9 cents out of a $4.19 payoff — about 2.1% of what the winner earned, taken after the takeout already took its share. Now run the same 9 cents against a short-priced winner. A computed payoff of $2.39 pays $2.30; the kept 9 cents is about 3.8% of that payoff. Run it against a long-priced winner: a computed $40.19 pays $40.10, and the same 9 cents is about 0.2%. Breakage is a flat rounding per ticket, so it falls hardest, in percentage terms, on the shortest payoffs — on favorites, on place and show tickets, on the bets casual players make most. No single ticket loses much to it. A season of short-priced tickets pays it hundreds of times, always in the same direction. When this book speaks of an "all-in drag" of about 20% on win betting, it

means takeout plus breakage plus the ordinary slippage of prices moving after you commit — the full distance between the pool the public funded and the payoffs the public collected. The derivation below uses both the clean takeout rate and the all-in figure, so you can see which claim rests on which number.

Break-even: the one formula in this book

Every bet in this chapter reduces to one question: how often must this bet win, at the price offered, before it stops losing money? The question has an exact answer, and you already own the arithmetic.

A price quoted as odds — 2-to-1, 4-to-1 — states the profit on a winning bet relative to the stake. A $2 bet at 2-to-1 profits $4 and returns $6 in total, stake included. A $2 bet at 4-to-1 profits $8 and returns $10. Call that total return the **gross return**: stake plus profit. If a bet pays a gross return of G dollars on a $2 stake, and it wins a share p of the time, then over many bets the average return per bet is p × G. The bet breaks even when the average return equals the stake: p × G = $2, which gives the break-even win rate:

Break-even win rate = stake ÷ gross return = $2 ÷ G.

That is the whole formula. Apply it across the prices a casual player sees every day. Each line below also works the dollars over 100 bets — $200 staked, $2 at a time — so you can watch the formula and the money agree.

Price	$2 ticket returns	Break-even win rate	What 100 bets must produce
Even money (1-to-1)	$4.00	$2 ÷ $4.00 = 50.0%	50 winners × $4.00 = $200

Price	$2 ticket returns	Break-even win rate	What 100 bets must produce
3-to-2	$5.00	$2 ÷ $5.00 = 40.0%	40 winners × $5.00 = $200
2-to-1	$6.00	$2 ÷ $6.00 = 33.3%	33⅓ winners × $6.00 = $200
3-to-1	$8.00	$2 ÷ $8.00 = 25.0%	25 winners × $8.00 = $200
4-to-1	$10.00	$2 ÷ $10.00 = 20.0%	20 winners × $10.00 = $200
6-to-1	$14.00	$2 ÷ $14.00 = 14.3%	about 14 winners × $14.00 = $200
9-to-1	$20.00	$2 ÷ $20.00 = 10.0%	10 winners × $20.00 = $200

Read the table from the price's side and it says something casual players feel but rarely state: the board is quoting you a required accuracy. Even money does not mean "this horse is a good bet." It means "you must be right half the time at this price, or you lose." A 4-to-1 price does not mean "this horse is a bad bet." It means "you must be right more than once in five, or you lose." The horse's quality enters only through the comparison between the accuracy you can actually deliver and the accuracy the price demands. Most losing at the track is not bad picking. It is delivering 45% accuracy to a price that demanded 50%, or 17% accuracy to a price that demanded 20%, race after race, without ever writing the two numbers side by side.

Worked example 1: even money, in dollars

You bet $2 to win, 100 times, on horses that go off at even money. You stake $200. Each winner returns $4.00.

Break even and you need 50 winners: 50 × $4.00 = $200 back on $200 staked. Now grade three handicappers against that bar. The first wins 50 of the 100. Return $200, profit $0 — a skilled picker of even-money horses who has worked a season for nothing, because the price demanded exactly what he delivered. The second wins 45: 45 × $4.00 = $180 back, a loss of $20, a return of −10%. He picked the winner nearly half the time, on the shortest-priced horses on the card, and lost a tenth of his money. The third wins 40: 40 × $4.00 = $160 back, a loss of $40, a return of −20%. That third record — win 40% of the time at even money — describes a player who is right more often than most players are right about anything, and it loses at the same rate the crowd loses on favorites. Accuracy is not money until the price says so.

Worked example 2: 4-to-1, in dollars

Same stake: $2 to win, 100 times, $200 in, on horses at 4-to-1. Each winner returns $10.00. Break-even needs 20 winners: 20 × $10.00 = $200.

Win 16 of the 100 — a record that feels cold, four losing weeks for every good one — and you collect 16 × $10.00 = $160, a loss of $40, the same −20%. Win 20 and you stand exactly still. Win 25 of the 100 and you collect 25 × $10.00 = $250, a gain of $50, a return of +25% — at a win rate, 25%, that the even-money player above would call a slump. The two examples are the same $200 and the same formula. The difference is entirely in what the price demanded: 50% accuracy at even money, 20% at 4-to-1. This is why this book never reports a win rate without a return beside it, and why you should never accept one without the

other from anyone else. A win rate is a claim about picking. A return is a claim about picking *at prices*. Only the second one spends.

The crowd-matching machine, and the 25% arithmetic

Now the derivation Chapter 1 asserted and this chapter owes you in full: why a bettor who merely matches the crowd loses the takeout, and how much better than the crowd you must be to break even in a win pool.

Imagine your money flowing through win pools for a season: $2,000 staked, $2 at a time, across many races. Suppose your picking matches the crowd's exactly — you hold, in effect, a representative slice of the public's own tickets, at the public's own prices. The pools you bet into remove their takeout before paying. At a 16% takeout with no other drag, your $2,000 of representative tickets can only claim its share of what the pools paid out, and the pools paid out 84 cents per dollar. Your season returns 0.84 × $2,000 = $1,680. You lose $320 — exactly the takeout on your flow, 16%. Matching the crowd is not a neutral result. It is a machine that converts your stake into the track's fee at the pool's rate.

To break even, your tickets must return $2,000 instead of $1,680. The required improvement is a ratio, not a feeling: $2,000 ÷ $1,680 = 1.1905. At a clean 16% takeout, you must do about **19% better than the crowd in money terms** — collect about 19% more return from the same stake at the same prices, which in practice means naming about 19% more winners at the same

average payoff, or the same winners at prices about 19% richer than the crowd's price for its own accuracy, or some mix of the two.

Now use the all-in drag instead of the clean rate. Takeout at 17%, plus breakage falling on your winning tickets, plus the small losses from prices that move against you between your decision and the close, puts the working drag on a typical win-pool season at about 20% — the figure behind Chapter 1's statement that blind favorite money loses about a fifth. Run the same machine at 20%: your $2,000 returns 0.80 × $2,000 = $1,600, a loss of $400. Break-even now requires $2,000 ÷ $1,600 = 1.25. **Twenty-five percent better than the crowd.** That is the arithmetic, shown: at an all-in drag of one fifth, a crowd-matching bettor keeps four fifths of his money, and climbing from four fifths back to whole requires a gain of one quarter on what he kept — 0.25 × $1,600 = $400, exactly the loss. "Roughly 25% better" is not a slogan about effort. It is the reciprocal of 0.80.

Count it in winners once, to fix the scale in your head. Bet $2 to win 100 times on horses whose winners pay an average gross return of $5.00 — a 3-to-2 average price, chosen to make the counting clean. You stake $200. At the crowd's accuracy on horses it prices this way, you win 32 of the 100: 32 × $5.00 = $160 back, a loss of $40, the 20% drag in dollar form. Break-even at the same average price needs $200 ÷ $5.00 = 40 winners. Forty winners against the crowd's 32 is 40 ÷ 32 = 1.25 — 25% more winners, at the same price, sustained over the whole season. Sit with that number the next time a tip sheet, a television analyst, or your own hot afternoon tells you your reading of the races is a little sharper than the room's. A little sharper is 5%. The fee

demands 25. The gap between those two numbers is where almost all handicapping confidence goes to die, and it is why the lab in this book treats a measured edge of a few points of return as a real result instead of a disappointment. Against a 25% requirement, small true edges are the only kind anyone has ever documented at full counts.

The same reciprocal gives you the requirement in any pool. Divide 1 by what the pool returns per dollar. A win pool at a clean 16% returns 0.84, and $1 \div 0.84 \approx 1.19$. An exotic pool at 24% returns 0.76, and $1 \div 0.76 \approx 1.32$ — about **32% better than the crowd** in money terms, in that pool, before you have earned a cent. Count the exotic season in dollars to match: \$2,000 of crowd-matching exacta money returns $0.76 \times \$2{,}000 = \$1{,}520$ and loses \$480, against the win pool's \$1,680 and \$320 on the same stake and the same quality of opinion. The exacta bettor did not handicap worse. He paid \$160 more in fees for the privilege of expressing the same opinion in a more expensive room.

Place and show: the same fee on a smaller prize

Straight pools below win betting run on the same machinery with one twist: the net pool splits. In a place pool, after takeout, the money divides into two equal parts — one part for the tickets on the horse who finished first, one part for the tickets on the horse who finished second — and each part is divided among that horse's place tickets. A show pool splits three ways. The fee comes off the top exactly as in the win pool. What changes is the size of each slice.

Work one place pool. Stated numbers: $6,000 bet to place, takeout 17%. The track removes 0.17 × $6,000 = $1,020. The net pool is $4,980, split into two halves of $2,490. Suppose $1,000 was bet to place on the winner and $500 to place on the runner-up. The winner's place tickets claim $2,490 ÷ $1,000 = $2.49 per dollar, so a $2 place ticket on the winner returns $4.98 before breakage — $4.90 after dime breakage takes its rounding. The runner-up's place tickets claim $2,490 ÷ $500 = $4.98 per dollar, so a $2 ticket returns $9.96, paid as $9.90. Two lessons hide in that small table. First, your place payoff depends on which other horse shares the finish with yours, because each finisher's backers divide a fixed half — a heavily bet co-finisher dilutes nothing of yours, but a heavily bet *your* horse dilutes every ticket on it. Second, the break-even formula from the win pool applies unchanged, price by price: a place ticket that returns $4.90 must cash more than $2 ÷ $4.90 ≈ 41% of the time to earn its keep. Finishing in the first two is easier than winning, and the price knows it, and the fee takes the same cut of the easier bet. There is no pool in the building where the arithmetic is suspended.

The same opinion in two pools: a decision, worked

Put the machinery together on one handicapping opinion and watch it produce two different decisions. You read a race and land on a horse. Your honest estimate — stated as a number, the way Chapter 19 will teach you to state it — is that the horse wins 30 times in 100 races like this one. The win board offers 5-to-2: a $2 ticket returns $7.00, and the break-even rate at that price is $2 ÷ $7.00 ≈ 28.6%. Your estimate, 30%, clears the bar, narrowly. Bet $2 to win 100 times: you stake $200, you win 30 times, you collect 30 × $7.00 = $210. Gain: $10, a return of +5% —

before breakage and price movement shave it, which is why a narrow paper edge is a reason for caution rather than celebration. The opinion, in this pool, at this price, is a small positive bet.

Now express the same opinion as an exacta: your horse on top of the one specific rival you expect to finish second. Your honest estimate for that exact pair, in that exact order, is 8 times in 100. The exacta pays 10-to-1 on the board: a $2 ticket returns $22.00, and the break-even rate is $2 ÷ $22.00 ≈ 9.1%. Your estimate, 8%, does not clear the bar. Bet it 100 times anyway: $200 staked, 8 hits, 8 × $22.00 = $176 collected. Loss: $24, a return of −12%. Same race. Same reading of the horses. Same confidence, honestly stated both times. One expression of the opinion earns $10 and the other loses $24, and the difference is not luck — it is the price demanded against the chance delivered, in a pool that also removed a larger share before paying. When a player tells you he "had the winner" in a race where his exacta lost money, both halves of his sentence are true, and only the second half is about betting. Picking the horse and pricing the ticket are separate skills. Takeout taxes the second one.

The reverse case exists and matters just as much. Prices in a large exotic pool are set by the money spread across hundreds or thousands of combinations, and a specific combination can be priced richer than its true chance while the win price on the same horse is priced tight. The discipline is not "win pools good, exotics bad." The discipline is mechanical: for any pool you are considering, state your estimated chance, compute the break-even rate from the price that pool actually offers, and compare the two numbers before the money moves. The pool

with the lower takeout gives your opinion more room to be slightly wrong. The pool with the higher takeout punishes the same honesty harder. That is the entire practical content of this chapter, and it fits in the margin of a program.

What takeout does not do

Three clarifications keep the fee in proportion, because players overcorrect on it as often as they ignore it.

Takeout does not make every bet equal. The fee is a percentage of the pool, removed before payoffs; it does not reach into individual bets and flatten them. A mispriced horse — one whose true chance clearly exceeds the chance its price implies, by more than the fee — is a profitable bet in expectation, in this chapter's arithmetic as much as anywhere. The fee sets the size of the mispricing you need. It does not say mispricings never reach that size. Chapter 1's record says they rarely do, and that the crowd's prices are usually close. Rarely is not never, and the chapters ahead are a map of where "rarely" concentrates.

Takeout does not reward volume. Betting more races does not dilute the fee; it multiplies the number of times you pay it. A player who bets $2,000 a season at a 20% all-in drag while matching the crowd loses about $400 whether the stake arrived as 100 bets or 1,000. The 1,000-bet version merely pays it in smaller installments. If anything in your routine feels like it is working because you are betting more, check whether it is working or just paying more slowly.

Takeout does not care about your staking plan. Chapter 18 presents the lab's staking and stopping guidelines for what they are — tools that manage the size and speed of losses. A staking

plan decides how much of your money is exposed to the fee on each race. It cannot change what the fee does to each dollar exposed. No arrangement of bet sizes turns the $1,680 season above into $2,000. Only accuracy beyond the crowd's, at prices that pay for it, moves that number, and it moves it by the small amounts full counts support.

The three lessons of the fee

First: the fee comes out before you are paid. In a $10,000 win pool at 16%, the winners divide $8,400. In the same pool as exactas at 24%, they divide $7,600. Every payoff you will ever collect is a share of what is left after the removal, and breakage rounds your share down again — hardest, in percentage terms, on the shortest prices.

Second: every price quotes you a required accuracy. Break-even win rate = stake ÷ gross return. Even money demands 50%. Four-to-one demands 20%. State your honest chance, compare it with the demand, and let the comparison — not the horse, not the story — decide the bet.

Third: beating the crowd a little is losing. A crowd-matching win-pool season at a 20% all-in drag returns $1,600 on $2,000. Break-even requires $2,000 ÷ $1,600 = 1.25 — roughly 25% better than the crowd in money terms — and the same reciprocal in a 24% exotic pool requires about 32% better. That requirement is the opponent in every chapter that follows. Methods earn their place in this book by showing, on full counts, that they close some real part of that gap. None of them repeal it.

Exercise: price your last twenty bets

Take your last twenty bets — from your records if you keep them, from memory and ticket stubs if you do not, and let the gaps in the record teach their own lesson about Chapter 18. For each bet, write three numbers: the pool (win, place, exacta, and so on), the gross return the winning ticket would have paid on $2, and whether the bet won.

Compute two rates for the twenty. First, your actual win rate: winners ÷ 20. Second, your break-even rate: add the gross returns as if each bet's price applied to all twenty — or, more simply and honestly for a mixed set, compute what the twenty bets actually returned against the $40 they cost, and set that beside the win rate. Then do the exercise's real work: for each bet, write the chance you honestly believed, before the race, that the bet would cash — and if you cannot reconstruct that number for most of the twenty, notice that too, because it means the comparison this chapter runs on was never available to you at the window.

Most players who do this find the same shape the pools predict: a handful of bets whose price demanded far more accuracy than any honest estimate could have supplied, a middle that roughly matched, and one or two bets where the price was genuinely generous. The twenty bets will not prove anything — Chapter 4 will explain why twenty of anything proves very little. But the habit will: state the chance, compute the demand, compare before you bet. Takeout is a tax that decides everything only for players who never do that arithmetic. For players who do it, the tax is still collected on every pool — it just stops collecting the extra, voluntary surcharge of bets whose prices never gave them a chance.

Chapter 4 — Finding is not proving

Every handicapper has had the same evening. You spread the charts across the table, you look at the day's winners, and a pattern rises off the page. Horses that dropped in class won. Horses making their third start off a layoff won. Horses whose rider changed to the leading apprentice won twice at prices. By midnight you hold a rule. It fits the day perfectly, because you built it out of the day.

The next Saturday, the rule loses.

This chapter is about why that happens, and about the discipline this lab built to keep it from happening to us in print. The discipline has a plain name. Finding a pattern in races you have already seen is *finding*. Proving is a different act: you write the rule down, in full, before the races that will judge it, and you let races nobody tuned the rule on decide. Most of what is sold as handicapping knowledge is finding dressed as proving. The difference between the two is the difference between this book and a tip sheet.

Three ways a true-looking result can be false

A backtest — a rule run over old races — can lie to you in three distinct ways. They have technical names. Learn them in plain language, because every later chapter in this book leans on them.

Selection bias: you kept the winners of your search. Suppose you test fifty ideas on the same year of races. Forty-nine lose. One wins, by a margin that looks real. If you publish the one

and bury the forty-nine, your reader sees a discovery. What actually happened is a lottery: test enough losing ideas on the same races and chance alone promotes one of them to the top of the pile, the way one ticket in a drum has to win. The races did not certify the idea. Your search selected it, and the selection — not the horses — produced the shine. The honest unit of evidence is never the winning idea alone. It is the winning idea *plus the count of ideas that were tried and discarded to find it.*

Overfitting: the rule memorized the test. A rule has knobs: thresholds, weights, exclusions. Turn the knobs while watching the results and the rule gradually stops describing horse racing and starts describing *those* races — the particular wet Tuesday at one track, the one longshot whose payoff carries the whole year, the three weeks when a single barn was hot. The fit improves with every adjustment, and the improvement is worthless, because the races doing the judging are the same races that did the teaching. A student who copies the answers into the exam booklet scores perfectly. The score measures the copying, not the student.

Multiple testing: the same trap, counted. Run one test at the usual standard of evidence and a false alarm is uncommon. Run a hundred tests at that standard and false alarms become a schedule, not a risk — you expect several, on pure chance, even if every idea you tested is worthless. Handicapping research runs hundreds of implicit tests: every angle in every book, every split of the data by surface, distance, class, and price band. Split a result enough ways and one slice will glow. The glow is the arithmetic of slicing, not a fact about horses.

Notice what the three traps have in common. None of them requires a dishonest researcher. They punish honest, hardworking counting, done in the wrong order. The order is the whole game: rule first, races second — or it did not happen.

A word about where these traps hide in ordinary practice, because they rarely look like statistics. Selection bias is the notebook with the losing months torn out — not from dishonesty, but because the player stopped recording when the system "wasn't being played seriously anymore." Overfitting is the angle that gains a condition every time it loses: it works on turf, then on turf sprints, then on turf sprints after a layoff with a rider change, each condition added to rescue the rule from a defeat, until the rule describes eleven historical races precisely and the future not at all. Multiple testing is the Saturday spent scanning forty charts for "what won today," a search guaranteed to return a pattern because forty charts contain more patterns than any one day can repeat. And the mildest form of all is the holdout you peek at: the player who "just checks" how his new rule would have done last month, adjusts one threshold, checks again, and after six checks owns a rule fitted to last month in every detail while sincerely believing it was tested there. None of these players is fabricating. Each did real counting. The counting was simply arranged so that it could only agree.

What "out of sample" actually means

The fix has a name you will see in print everywhere, usually used loosely: *out of sample*. Here is the strict meaning this lab uses. A result is out of sample only if the races that produced it played no part — none — in choosing the rule, setting its thresholds, picking its features, or deciding which version of it

to report. A holdout set you peeked at six times while tuning is not a holdout. It is a second training set with better public relations.

There is a subtler leak, and it cost this lab its most promising model. A feature — a number fed to a rule — can itself carry information from the future. Build a table from a full year of results, then use that table to "predict" a race in January of that same year, and the January prediction quietly contains December's outcomes. Nobody cheated. Nobody even made an error in arithmetic. The information simply flowed backwards in time through a shared table, and the backtest credited the model with knowledge no bettor at the January window could have held. We call that *point-in-time leakage*, and the next section tells the story in full, because it is the most instructive failure in our records.

First, the standard the story produced. The lab now builds its research data under a rule it calls Clause 7, adopted after the audit you are about to read: **every per-starter feature must be constructible from records dated strictly before its target race.** A feature that cannot meet the standard must be declared, and every headline computed with it must be printed next to a sensitivity — the same headline recomputed with the feature rebuilt honestly or removed. Clause 7 is not a courtesy. It is the difference between a measurement and a memory.

The model that knew too much: LOGIT_ALL

In the autumn of 2026 the lab ran a ranking search on its corrected 2023 research corpus. One candidate stood out. Call it LOGIT_ALL: a logistic regression — a standard statistical

model that turns a set of numbers into a ranking — over eight pre-race features: the morning line's implied probability, our own speed figure, our handicap weight, early and late pace percentiles, a draw index, days since the last race, and prior starts. It was fitted on the first half of the year and then scored once, untouched, on the second half, its holdout.

On that holdout it did what no other candidate in the search did. It beat the morning-line favorite head-to-head in the duel, in the races where its top pick differed from the favorite. Betting on its top pick in every holdout race lost money — it lost money, as everything in Chapter 1 loses money — but the favorite lost more, so the paired difference was positive in the model's favor.

Read that result the way this book has taught you to read it. The model still failed both of the search's frozen gates. The BETTER RANKING gate required a paired margin of at least +3.00 points; the model posted a positive margin that fell short of that gate. The PROFIT gate required a positive return; the model posted a negative return. It was, on its own record, a better ranker than the favorite that lost less money than the favorite. The lab locked it — froze its coefficients in a file, with a scorer that refits nothing — as a *paper candidate record only*. It files no picks. It was never adopted. Then, before a prospective judge built to watch it on future races was trusted, the lab audited it forensically, and invited outside reviewers to audit the record too.

An outside reviewer named the problem first. One of the eight features, the draw index, was suspect on point-in-time grounds. The forensic audit then did the work: it traced every feature to its source records and tested each one for dependence on information dated on or after the race it described.

The draw index failed the trace. The index assigns each post position a value drawn from a table — track, distance band, surface, field-size band — and the frozen table had been built from the *full* 2023 charts corpus. A horse running in January was scored with a draw value computed partly from races run the following autumn. The model's other seven features survived the audit; this one channel carried the future inside it. Two other suspects were cleared on evidence, not on reassurance. The identity check hunted for horses whose speed figure had been built from another, same-named horse's races: it found a small number of ambiguous starters, a tiny fraction of all starters, and neutralizing their figures did not move the holdout result. The leak was not everywhere. It was in one pipe, and the audit's job was to measure what that pipe carried.

The measurement is called a sensitivity, and it is the honest way to answer the only question that matters: *how much of the edge survives without the leaked information?* The auditors rebuilt the draw table as an expanding window — each race scored only from charts dated before it — and re-ran the holdout with the model's frozen coefficients unchanged. The paired edge fell sharply, to roughly half its stored size. The duel net fell sharply as well. Of the surviving edge, the reconciliation of the audit round attributes most of it to the draw channel even in its honest, causal form; silence the draw feature entirely and the paired edge falls to near zero, with an uncertainty interval that

crosses zero. The strictly causal figure carries a bootstrap confidence interval that stays positive, but at roughly half the size the stored record advertised, and nowhere near the ranking gate the search required.

Sit with the shape of that result, because it is the chapter in miniature. The leak did not manufacture the whole edge out of nothing; a smaller edge survived the repair. What the leak manufactured was the *headline* — the stored paired figure that made the model look like the search's one success. And notice what the lab's record now licenses anyone to say, in the audit round's own binding wording: the model reproduces exactly, its features are faithful except for the one declared defect, its point-in-time edge is smaller and draw-carried, it failed its gates, and it was never adopted. Quoting the stored headline without the restatement is, in the lab's vocabulary, a citation defect. That is what proof discipline looks like from the inside: it does not just correct the number. It governs how the record may be repeated.

Two lessons generalize beyond one model. First, leakage rarely announces itself. The draw table was not a suspicious input; it was a frozen reference table, the kind of background furniture every handicapping system uses — post-position statistics, track pars, trainer tables. Any statistic built from a period that includes the race being predicted is a pipe from the future, and the pipe does not care that you built it innocently. Second, the defense is procedural, not vigilance-based. You cannot promise to be careful enough. You freeze the construction rule — Clause 7 — and you audit against it, with the verdict rules written before the audit runs. The LOGIT_ALL audit's own verdict rules were frozen in advance: integrity fails if any feature depends on a record dated on or after its race, if ambiguous

identities exceed half a percent of starters, or if the point-in-time sensitivity collapses the edge below its pre-set floor or the duel to zero or below. The gates did not move after the fact, in either direction. A stricter bar proposed during the round — higher than any frozen rule the lab had — was recorded as advice and not adopted retroactively. Moving the bar after you see the result is overfitting applied to judgment itself.

The family that reversed: FAIR thresholds

The second episode is quieter and, for the working handicapper, more common. It involves no model at all — just a threshold, the kind of rule a player can hold in his head.

The idea is an old one. Estimate each horse's chance of winning from the available evidence. Compare the estimate to the price. Bet only when the estimate exceeds what the price implies by a wide margin — in the lab's test, when the estimated probability times the payoff multiplier reached a threshold of 1.50, a fifty-percent cushion over fair. Four betting arms were built on that trigger: bet the favorite only when it qualifies; bet every qualifier flat; split the stake across qualifiers in proportion to their estimates; bet the top-ranked qualifier. Before any arm could be called a pass, it had to clear a bundle of conditions fixed in advance: a positive return, both halves of the year above –2%, a minimum number of bets, a limit on how much of the result could come from a single month, and a win over the identical rule run on morning-line probabilities alone by at least 3 points. That last condition — the ML-only control — deserves your attention. It asks the killing question: *does your probability estimate add anything to the estimate already printed in the program?* A fancy method that cannot beat its own control is decoration.

On the lab's earlier, uncorrected data, arms of this family had produced attractive numbers. Then the corpus was rebuilt — records corrected, point-in-time discipline applied — and the family was re-run on the corrected research corpus. The verdict in the record is two words: **ALL REVERSED.** No arm passed the bundle. Every arm failed the return leg and both half-year legs. When the battery was re-run on the second research population, nothing flipped back: every arm lost money and failed again. The verdict was NOT A CLEAN PASS for every arm. One curiosity in the record makes the chapter's point by itself: the favorite-only arm's ML-only control — the same trigger, using nothing but the morning line's own probabilities — returned a positive figure. On a very small number of bets. A very small number of bets is not a discovery; it is a coin that landed kindly. The control's thin count changed nothing, and the arm failed on both populations regardless. But keep the image: the only positive number in the whole family test sits on the smallest denominator in the room. That is where false findings live — in the slices too small to average out their luck.

Why did the family reverse? Not because the threshold idea is foolish; a price cushion is a sensible demand, and Chapter 5 will show the arithmetic that makes it sensible. It reversed because the earlier numbers had been computed on records that did not survive correction, and because a result assembled from many arms, many splits, and many small slices will always contain a few bright cells. Test a family, and the family's *best* cell is a selected statistic — selection bias again, wearing a lab coat. The discipline that caught it is unglamorous: a pass bundle written in advance, a control that uses only public information, both halves of the year required to stand up separately, and a

concentration limit that refuses to let one month carry the verdict. Each condition closes one of the three traps from the start of this chapter.

Why the lab freezes rules

You can now see the reason for the practice this book keeps mentioning: the lab writes its tests down before it runs them, and freezes its rules before forward races are allowed to count. The written-down-in-advance document — the preregistration — states the rule in full, the population, the exclusions, the comparison, and the gates, with numbers. It exists because every researcher, this lab included, can feel the pull of the knobs after the results arrive. A threshold at 1.50 that fails becomes a threshold at 1.65 that "was always the intent." An exclusion added after a bad month becomes "obviously correct." Each edit is small, defensible, and fatal: the test quietly becomes a fit, and the result becomes a finding instead of a proof. The freeze is a promise made to the one person you cannot fool later, because you wrote down what you believed earlier.

The amendments the lab adopted after these episodes read, in plain terms, like scar tissue, and the lab keeps them attached to every verdict:

- **Scope.** Every verdict carries its corpus and universe wherever it is cited — the population label is part of the result, not a footnote. A verdict shown to differ between two corpora on the same test goes back to a human decision automatically. The FAIR family is the reason: a verdict without its population is a rumor.

- **Two grades.** A method can be adopted as a *pick rule* — evidence that it orders horses usefully — without any wagering meaning. It earns wagering meaning only when its economics interval clears zero. Chapter 1's consensus record holds the first grade in its research tests and not the second. The two grades keep a true sentence ("it ranks better") from smuggling in a false one ("so bet it").
- **Concentration.** Before the gates, every candidate reports how much of its edge comes from its largest factor, cell, track, and price band. An edge that lives in one place is presumed to be a spike — noise — until the concentration is explained. You met this in Chapter 1: most of one method's net came from one price band and one half of the year.
- **Displacement.** Any rule that overrides another rule ships with a standing watch on the horses it displaced and a numeric tripwire written in advance. Overrides are where tuned systems hide their edits; the watch makes the edit pay rent in public.
- **Independence.** A proposed method must state what information it uses that the current method does not already contain. A re-weighting of the same factors is a variant of the old method, not a new one, and it does not get a new method's fresh hearing. Without this rule, the same finding can be "discovered" annually by rearranging it.

None of these amendments makes a bad idea win. They do something more valuable: they keep the lab's *record* honest, so that when this book tells you an idea was not supported, the words cost something — a frozen rule, a full count, a comparison named in advance, and a file a skeptic can re-run.

What this means at your kitchen table

You do not run a laboratory. You have a notebook, a program, and Saturday. The discipline scales down intact.

Write the rule before the races. All of it: which horses qualify, which races you skip, what price you require. A rule you cannot write in two sentences without the word "unless" is not a rule yet; it is a mood. Date the page. Then count *every* race the rule touches, including the embarrassing ones — the denominator is the discipline. Compare against the dumbest honest baseline, which in this sport is always the same: what happened if you simply bet the favorite, or followed the morning line's order, in those same races. If your rule cannot beat its control, you have learned something worth the price of the notebook. And when a result delights you, ask the auditor's three questions before you believe it. How many ideas did I try before this one? Did any number in my rule come from the races I am now using to judge it — a table, a par, a "this track favors closers" note built from the same season? And how many bets is this actually standing on?

Finding is the fun of handicapping, and this book does not ask you to give it up. Spread the charts. Chase the pattern. Just label the evening's work correctly: a hypothesis, found. Proof begins the next morning, with the rule frozen and the races unseen.

Everything after this chapter is graded by that standard — including, in the next chapter, the uncomfortable case of a method that proved it could rank horses better and still could not be bet.

Chapter 5 — Better ranking, worse bets

Chapter 1 ended with a record that should bother you a little. Our lab's consensus method — five factors scoring the morning line's top four, a price cap on top — carried a filed selection in 95 races on the lab's own settled product record: the filed cards of November 2–4, 2026 — 111 races. On the paired races in that small record it did not beat the favorite, and it lost money. This chapter takes that record apart slowly, because inside it sits the single most misunderstood fact in handicapping: **a method can get better at naming the order of finish and worse at making money, at the same time, for reasons that are pure arithmetic.** Once you see the arithmetic, you will read every claim in every handicapping book — this one included — with a new question ready: *better at what?*

Four claims, kept apart

This book grades four different claims, and this chapter is where the grading starts to matter.

An **association** claim says two things move together: favorites with no early-speed rival win more often than favorites crowded on the lead. It describes a population. It picks no horse and prices nothing.

A **ranking** claim says a method orders a field better than some comparison: put its first choice against the morning line's first choice over thousands of races and count who named more winners, or whose choice finished ahead head-to-head. Ranking is about order only. Money does not appear in it.

A **probability** claim says a method's stated chances match reality: of all the horses it calls 30% chances, about 30% win. That property is called *calibration*, and Chapter 5's second half returns to it. A method can rank beautifully and be badly calibrated — always too confident, always too timid — and a calibrated method can rank no better than the program.

A **returns** claim says the bets the method generates, at the prices actually offered, return more than they cost. This is the only claim your bankroll experiences. It is also the hardest to earn, because between a ranking and a return stand two toll collectors: the price, and the takeout.

None of the four claims implies the next. Each needs its own test, its own denominator, and its own verdict, and a method earns them one at a time or not at all. A book — or a tipster, or a figure maker — that shows you evidence for one claim while letting you assume the others is asking you to grade its exam on the questions it chose.

Most handicapping arguments are two people grading different claims and calling each other wrong. The pace handicapper proves his figures rank horses well — often true. The skeptic answers that pace bettors lose money — also often true. Both claims can hold in the same race on the same day. Keep the four claims labeled and most of the sport's literature becomes legible: a long parade of ranking evidence offered as if it were returns evidence.

The arithmetic that sits between order and money

Define the one quantity a bet cares about. A horse wins with some true probability — call it *p*. The board offers decimal odds — call them *d*, meaning a winning one-dollar bet returns *d* dollars in total, stake included. A 2-to-1 horse has d = 3. Your **expected return** per dollar is p × d. If p × d is greater than 1, the bet returns more than it costs, on average, over many repetitions. If it is less than 1, the bet loses on average no matter how often the horse wins. At p × d exactly 1, the price is *fair*: the bet is a coin you pay nothing to flip.

Everything in this chapter is a footnote to p × d.

Watch what it does to a good horse. A favorite wins 40% of its races — a strong favorite, the kind any ranking method is proud to name first. It goes off at 6-to-5, d = 2.2. Expected return: 0.40 × 2.2 = 0.88. You lose 12 cents per dollar on the best horse in the race, ranked correctly, winning two times in five. There is no error in the ranking. The price simply does not pay for the risk, because the crowd saw the same horse you did and priced it first. Now watch a worse horse at a better price. A rival wins 20% of the time and goes off at 6-to-1, d = 7. Expected return: 0.20 × 7 = 1.40. It loses four races in five and returns 40 cents per dollar on average. The ranking prefers the first horse. The betting slip prefers the second. They are answers to different questions, and only one of them is asked at the window.

Where does the gap come from? The takeout, working through the prices. In a win pool the track removes roughly 15 to 17 cents of every dollar before payoffs are set, so the prices on *all* horses are shaded down together: add up the crowd's implied probabilities across a field and they total well over 100%, the

excess being the takeout and breakage expressed as probability. The consequence is a wall, and it is the same wall Chapter 1 measured: beating the crowd's *order* earns you nothing by itself, because the crowd's order is already inside the prices. To clear $p \times d > 1$ you must disagree with the crowd's *probabilities*, in the right direction, by more than the takeout's margin. A ranking improvement that agrees with the crowd — that promotes the same short-priced horses the crowd already likes, only more confidently — runs straight into the wall. The better you rank the obvious horses, the more faithfully you buy their shaded prices.

This is why ranking gains concentrate exactly where they cannot pay. Think about where any sane method can improve on the morning line's order. Not at the bottom of the field — nobody's refinement finds the 40-to-1 shot reliably. The improvements live at the top: choosing between the favorite and the second choice, promoting the third choice over a vulnerable favorite. Those are the shortest-priced horses in the race. Their prices carry the least slack, because the crowd prices them with the most money and the most attention. A method that picks the winner a little more often *among contenders* sounds like a breakthrough until you price the extra winners: they paid short prices. Meanwhile its mistakes — the favorites it talked itself off — were also priced short when *they* won. The duel can favor the method and the money can favor the crowd, because the duel counts wins and the money counts payoffs.

The break-even line, in dollars

Make the wall concrete. Take a win pool with a takeout in the typical range Chapter 1 described. The crowd's prices, taken together, pay back roughly 83 to 85 cents per dollar wagered across the whole field — the takeout arithmetic of Chapter 3, not a lab measurement. If your ranking of the field is exactly the crowd's ranking, and you bet your top choice every race, your long-run return is the crowd's return on favorites: a loss, after takeout. To reach a return of zero — not profit, zero — your top choice must outperform the favorite's money record by roughly a quarter, as Chapter 3 derived. In probability terms: at the prices favorites actually pay, you need to name winners several points of win percentage more often than the favorite does, *at the same average price*, or name the same winners at materially better prices by opposing the favorite where its price is worst. Almost every real method that survives testing does some of the second and little of the first, for the reason Chapter 1 gave: the crowd's win-rate ranking is the pooled work of everyone in the pool, and out-naming it by a wide margin is not a thing that happens. Out-pricing it at the margins — declining its worst-priced favorites, collecting a point or two of payoff the crowd left on the table — is the whole of the achievable game, and it is worth a few points of return, not twenty.

This is also why the short end of the odds board is a trap for improvers. Suppose your work genuinely adds information — you are right, measurably more often than the crowd, about which of the two favorites is stronger. Your reward is capped by the price: the stronger favorite pays a short price whether you found it by a season of figure work or by reading its name off the tote board. Information about short-priced horses is the

most competed-for information in the pool, so it is the most completely priced. The same unit of genuine insight applied where the crowd is thin — mid-priced horses, odd field shapes, races the public prices lazily — buys more expected return per insight, because the price has room to be wrong. A ranking method does not get to choose where its edge pays. The market chooses, and it pays least exactly where ranking is easiest.

The worked example: our own record, taken apart

Now the lab's record, with the pieces labeled by claim. The population is the lab's own settled product record: the filed cards of November 2–4, 2026 — 111 races. That is three days. The sample is small, the record is days old, and the uncertainty around every rate here is wide. Three days prove nothing either way. The record is published nightly as it grows.

The §26 selection book — a $2 win bet on the consensus selection in every race where one was filed, with a scratched selection treated as a refund, not a bet — carried 95 races with a filed selection and placed 90 bets. It won 25 of them, a 27.8% strike rate, staked $180.00, returned $101.20, and returned –43.78%.

The official-pick book — one official leg per race, with its filed backup chain when the pick scratches — carried 95 legs. It won 27 of them, a 28.4% strike rate, staked $190.00, returned $118.20, and returned –37.79%.

The headline book as published nightly — official legs plus the §16 counter-move legs — carried 108 legs. It won 28 of them, staked $216.00, returned $133.40, and returned –38.24%.

The consensus step changed the official pick in 12 of the 95 races. The price cap then capped 5 of those changed picks back to the favorite.

The ranking claim, against the official pick. The selection is chosen by a consensus vote; the lab also files an official pick by a separate rule, which gives the record a built-in head-to-head. On the small rebuilt record the two books stand close: the selection book returned worse than the official book, and won slightly less often. Three days cannot grade the committee against the single rule. In the lab's larger research tests the committee held a genuine duel and paired advantage over the official pick — a ranking finding reported here qualitatively, with its corpus numbers struck. As ranking evidence in research it was genuine: the committee named the better horse more often than the single rule did, and lost less money per dollar while doing it. The rebuilt filed record is too small to confirm or refute that finding.

The ranking claim, against the favorite. In the 85 races where both the selection and the morning-line favorite started, the favorite won more often — 25 winners, a 29.4% strike rate, against the selection's 24 winners and 28.2%. The money ran the same way on this small record. The favorite returned $103.90 on $170 staked, a return of −38.88%. The selection returned $99.00 on the same stake, a return of −41.76%. The paired difference is −2.88 points — on this small a record the selection did not beat the favorite. Let that stand in plain view: the crowd's first choice out-won our method's first choice and out-returned it, on three days of filed cards. In the lab's research tests the paired money comparison ran the other way, earned while naming fewer winners, because the selection's winners paid more on

average than the favorite's winners. That research finding does not transfer to the filed record by assertion. The filed record says what it says, and it is published nightly as it grows. Two scoreboards can have two winners over different populations and different periods. On the population this book prints, the favorite leads on both.

The returns claim, tested separately. A paired difference on a backtest is not yet a returns claim; Chapter 4 explained why. The lab therefore ran a separate economics test on an earlier, four-factor form of the consensus — a different pick object from the production selection, and the chapter keeps them distinct, as the lab's record does. Its headline paired point estimate was positive. Then the discipline bit. The bootstrap — resampling the races many times to ask how much of that edge is luck's width — put a confidence interval around it that crossed zero. In a substantial share of the resamples the edge was zero or worse. An interval that includes zero cannot support a returns claim; it can support a *watch*. The lab's gates required the headline to clear its pre-set margin, the interval's lower bound to clear zero, and no single price band or half-year to carry more than its pre-set share of the net. Every leg failed: the headline itself came in under its margin, the interval crossed zero, one price band carried most of the net, and the second half of the year carried more than the whole net — which means the first half, on its own, was negative. The verdict, in the lab's two-grade vocabulary: **ANNOTATION ONLY.** A pick rule with evidence behind it. No wagering meaning. Both halves of that sentence are the finding.

The disagreement races, up close. The economics test's internals show the mechanism of this whole chapter in qualitative form. In the races where the method disagreed with the favorite, the favorite won more often than the method's pick — the method's horse lost the duel outright on this pick object. But when the method's pick won, it paid more on average than the favorite paid when it won. The method was buying longer prices and collecting less often at more per collection, and the two effects nearly canceled: both sides lost money in those races, the method's picks by a little less. Nearly canceled is the honest summary of almost every real handicapping edge ever measured at this scale. The takeout takes its cut from both sides of the argument, and what survives is a difference of a few points, wide error bars, and a verdict that must be written in pencil.

Where the edge lived. The concentration report — the check Chapter 4 described — shows how little of the track the research edge was standing on. The band of races where the favorite's morning line sat between 3.5 and 6 supplied most of the entire net difference, and it did so on a small subset of the disagreement races, inside which the paired margin ran strongly positive while the other bands ran near zero or negative. Split the year instead and the same thinness appears: the first half's paired result was negative; the second half's was positive, which is why the second half accounts for more than the whole net. An edge that lives in a small subset of races and one half of one year may still be real — but it has the exact shape Chapter 4 taught you to distrust, and it cannot be told apart from a spike without new races judged under the frozen

rule. That, in one paragraph, is why the lab's verdict was annotation only rather than either celebration or burial: the evidence supported watching the method, and nothing stronger.

One more lesson from that test belongs in every reader's pocket. Spread the headline across *all* primary races — including the majority where the method simply agreed with the favorite — and the paired edge shrinks to a fraction of the disagreement-only headline. Headlines computed only on disagreement races describe a part-time method: in most races it is the favorite, wearing a lab coat. Whenever a system's literature quotes its record "when it disagrees with the public," ask for the all-races number too. Both are true. Only one of them is the method.

Calibration: the third claim, and the quietest failure

Between ranking and returns sits probability, and it fails quietly. A method is calibrated if its numbers mean what they say: gather every horse it rates a 25% chance, and about one in four should win. Calibration matters because the betting rule of Chapter 4 — bet when your estimate beats the price by a cushion — runs on estimates. Feed it inflated estimates and it manufactures "value" out of your own optimism. A model that calls 20% horses 30% will find overlays everywhere, bet them all, and lose at exactly the rate its calibration error predicts. Its ranking, through all of this, may be superb: the order can be perfect while every number attached to the order is wrong by a third.

This is not hypothetical bookkeeping. During the audit described in Chapter 4, the lab's locked ranking model was tested as a probability estimator as well as a ranker, and the

record's licensed wording carries the distinction explicitly: the model is a relative ranker only, not a probability estimator. Its order carried information; its implied percentages could not be spent. When you build your own estimates at the kitchen table, the same split applies. Sorting a field into your order is one skill. Writing a percentage next to each horse that survives contact with a season of results is a second, rarer skill, and the cushion in any value rule has to cover your calibration error before it covers the takeout. Players who skip the calibration question are not betting their ranking. They are betting their confidence, which is a different and much more expensive asset.

What to do with a ranking you trust

None of this makes ranking useless. It assigns ranking its correct job. A ranking you have tested — against the favorite, on full counts, under a frozen rule — is the best available answer to "who is most likely to win," and you need that answer before any price question can be asked. Use the ranking to build the short list: the horses whose prices deserve a second look. Then do the second, separate job: compare each short-listed horse's price to a probability you are willing to defend, and bet only where the price pays a cushion over your estimate — a cushion wide enough to cover your calibration error and the takeout together, which is wider than most players set it. Where no price clears the cushion, the ranking has still done its work: it told you the likeliest winner, and the prices told you the race was not worth a bet. Passing is not the ranking failing. Passing is the two claims being graded separately, in the same race, by the same player, which is the entire skill this chapter teaches.

Keep the records that let you grade yourself. For every bet, write down the horse, the price you took, and — before the race — your honest percentage. A hundred bets later you own three scoreboards: how often your top-ranked horse won (ranking), whether your percentages matched reality (calibration), and what the bets returned (returns). Most players keep only the third, in the form of a bankroll that shrinks without explaining itself. The three scoreboards together tell you which claim is failing, and they fail independently: a season of good rankings and bad prices looks identical, in a bare bankroll, to a season of bad rankings. The fix for the first is discipline at the window. The fix for the second is better handicapping. Confusing them is how players spend years improving the part of their game that was never broken.

The closing discipline

Here is the discipline this book applies from this chapter forward, stated as two sentences you can hold against any claim, ours included.

Ranking is a claim about order. A bet is a claim about price.

The lab grades them separately because they fail separately, and its own record is the standing example. On the lab's own settled product record: the filed cards of November 2–4, 2026 — 111 races, a small sample, three days, days old, with wide uncertainty — the consensus selection did not beat the favorite on the paired races: a paired difference of −2.88 points across the 85 races where both started, with fewer winners at worse returns. Its research economics reading held a positive headline whose uncertainty interval crossed zero and therefore proved no

wagering edge at all. The first sentence is the filed record. The second is a returns claim that did not clear its gate. Reporting either one as the other would be a small, fluent lie, and it is the lie most handicapping writing runs on: win rates quoted where returns were asked, paired edges quoted without their intervals, disagreement records quoted without the races where the method agreed and merely lost the takeout like everyone else.

At your own table the discipline costs nothing and pays in avoided bets. When a method impresses you — a figure, an angle, a sheet, this book's method — ask which claim its evidence supports. Order? Then use it to order, and let the prices decide the bets separately, with the favorite's record as the control and the takeout as the toll you must beat, not a detail you may skip. Price? Then demand the record in money: every race, the interval around the edge, and the halves of the year standing separately. A method that is honest about which claim it makes is already ahead of most of the literature. A player who insists on the distinction is ahead of most of the room — not because he ranks horses better, but because he has stopped paying ranking prices for betting decisions. That is the whole of Chapter 5, and it is load-bearing for everything after it: when the next chapters grade speed figures, pace, class, and consistency, each verdict will state its claim — association, ranking, probability, or returns — in its first lines, and mean it.

Chapter 6 — Speed figures: what they measure and what they hide

Open any past-performance sheet and your eye goes to the same place mine does: the big number. Each horse carries a speed figure for its recent races, and the sheet invites a simple reading. Bigger is faster. Faster wins. Sort the column, bet the top one, go home.

This chapter takes that number apart. A speed figure is not a measurement in the way a final time is a measurement. It is the end of a chain of decisions, and each decision in the chain can move the number by more than the gap between the horses you are trying to separate. The lab learned this in public, on the most famous horse in the country, when its own figure for a Breeders' Cup Classic winner landed in the middle of the pack and readers asked the obvious question: is your number broken?

The answer took days of work to establish, and it is the most useful answer in this book. The number was not broken. The number was doing exactly what it was built to do — and what it was built to do is narrower, and stranger, than the column on the sheet suggests.

What a figure actually is

Start with the one hard fact in a race: the clock. A horse runs a distance in a final time, and the beaten lengths behind the winner record how far back the rest finished. Everything after that is construction.

The first problem is that a final time means nothing by itself. A mile in 1:36 at one track, on one day, can be a strong performance; the same 1:36 at another track, or on another day at the same track, can be ordinary. Tracks differ — in surface depth, in layout, in how the turns bank. Days differ — wind, moisture, and maintenance move a surface up and down by amounts that dwarf the difference between good horses and average ones. A figure exists to strip those two layers away so that performances from different places and days can sit on one scale.

The stripping happens in two steps, and it helps to give them plain names.

Step one: the track constant. For each combination of track, distance, and surface, the figure-maker estimates how fast that course normally runs — in the lab's construction, a typical velocity built from every line in its store that matches the cell. A horse's raw figure is its performance measured against that constant: how much faster or slower than typical for that track, distance, and surface, with the beaten lengths of the other finishers converted the same way. On the lab's scale, zero is typical for the cell. A raw figure of +2 says the horse ran two points faster than the course usually produces. A raw figure of −2 says it ran two points slower. Raw figures have one virtue that matters later in this chapter: they depend only on the race itself and the constant. They do not change when the population around them changes.

Step two: the daily variant. The constant describes the course on an average day. Race days are not average. So the maker estimates how today's surface played — fast or slow relative to its own norm — and adjusts every figure from that card by that

amount. In the lab's construction this is called the projection variant, and the way it is estimated is the heart of the method: the chain solves each horse's established level from its prior races, then asks how today's whole field ran against those levels. If the horses on a card, as a group, ran faster than their established levels say they should, the day itself gets the credit — or the blame. The variant is that credit, in figure points, and it is subtracted from every raw figure on the card. A positive variant means the day ran fast, so each performance is cut. A negative variant means the day ran slow, so each performance is lifted.

Read that mechanism once more, because it is the sentence this chapter exists to teach: **each horse serves as its own control.** The variant does not come from a weather report or a groundskeeper's opinion. It comes from the horses — from the gap between what a card full of horses had already shown themselves to be and what they did today. When the field runs hot, the method concludes the track helped them, and it takes the help back out of their figures.

That is a reasonable design. It is also a design with a consequence you can watch in one famous case.

The Sovereignty file

On October 31, 2026, Sovereignty won the Breeders' Cup Classic at Keeneland. The chart records the win and a final time of 2:00.36 for the mile and a quarter. Another figure-maker's published list — the list that prompted the lab's exercise —

graded the performance at 136 on its scale, a number that sits near the top of that maker's all-time Classic list, alongside the names every fan knows.

The lab keeps its own figure for every horse in its records, built by the chain described above. After the race, it graded all 23 winners on that all-time list with its own chain, on one population: its production line store of 119,105 past-performance lines, merged with the 23 Classic cards parsed from the official charts — 121,816 lines in a single run, so that every winner's figure came out of the same solution. On the lab's scale, Sovereignty's Classic graded at +**0.81**. Of the 23 winners, that placed him **12th** — the middle of the pack.

The argument started immediately, and it is the argument every player has had about some figure on some sheet. One scale says the performance was near the top of a quarter-century of Classics. Another scale says it was ordinary by the standards of that company. Both cannot be describing the same thing — and that is the point. They are not.

Work the lab's arithmetic, because the file publishes its parts. The Classic card's dirt variant that day was +**2.04**: the chain judged that the Keeneland dirt surface on October 31 ran fast — the horses on the card, measured against their own established levels, ran about two points better than those levels predicted. Sovereignty's adjusted figure is his raw figure minus that variant. Working backwards from the two published numbers, his raw figure was +**2.85**: against the Keeneland constant alone, before any judgment about the day, he ran close to three points faster than typical. The variant then took two of those points away and handed them to the track.

So the whole argument fits in one line. The raw evidence says the performance was strong — a raw of +2.85 sits high in that company, on a scale where the largest raw the file prints for the whole table is Authentic's +3.45. The adjusted figure says: yes, and the day was helping everyone, so we do not credit the horse with all of it. Neither number is a lie. They answer different questions. The raw figure answers "how fast did he run against this course's normal?" The adjusted figure answers "how fast did he run, after we charge part of the performance to the conditions every horse on the card shared?"

Now look at what the same adjustment did to the rest of the table, because Sovereignty is not the extreme case — he is the teachable one.

The largest raw figure in the table belongs to Authentic, +3.45. His card's variant was +2.98, the largest positive adjustment in the table — his day ran very fast — so his adjusted figure is +0.47, and he lands 15th of 23. Flightline's raw was +1.96, the fifth-largest raw published in the table; a variant of +1.27 cut him to +0.69, 13th. The adjustment runs the other way just as hard. Saint Liam's raw was +0.01 — dead typical, nothing in it at all — but his card resolved a variant of −2.26, a slow day, and the lift carried him to +2.27 and 4th in the table. Arrogate sits at the top of the lab's list at +3.53, helped by a −3.33 variant on his day. The ordering of the adjusted table is, to a degree nobody expects before they see the parts, an ordering of the variants.

Hold on to the shape of that. In a table of 23 of the best performances of the modern era, the spread in raw figures between the famous names is a few points. The spread in variants is also a few points — from −3.33 to +2.98 in the cases above. **The adjustment is the same size as the thing being**

adjusted. That is not a defect in this one table. It is a property of figure-making everywhere, and it is the first thing a figure hides: the number on the sheet is a sum of a performance and a judgment, and the judgment is routinely as large as the performance gap you are trying to bet on.

The control check: did the figure at least predict him?

A figure has a second job in the lab's product, beyond grading the past. The figure printed for a horse before a race is a projection: the median of its adjusted figures from its last three or fewer scored lines. Before the 2026 Breeders' Cup, the lab ran its chain over both Keeneland cards and compared every starter's projection with the figure it actually earned on the day. That comparison is a control check on the whole construction, and its results keep the Sovereignty story honest.

Sovereignty's projection was +3.21 — the median of his last three adjusted lines. His actual figure on the day was +0.81. The difference, –2.40, is one of the larger misses on the card, and it came in a winning performance: he won the Classic while running below his own recent form as the chain measured it. The horse with a higher projection in his race, Knightsbridge at +3.34, earned +0.72 and finished third. The projection ranked the field; the race declined to follow the ranking in detail, as races do.

Across both cards, 114 of the 157 Breeders' Cup starters carried a projection at all — the other 43 had no scored line in the lab's records, a gap concentrated in the Friday juvenile turf races full of European shippers and lightly raced two-year-olds. For the

114, the average difference between actual and projected was –0.549 points, and the median was –0.442. That lean has a mechanical cause the report states plainly: both dirt cards drew large positive variants (+2.77 on Friday, +2.04 on Saturday), so dirt horses' actual figures were cut by the full variant, while their projections had been earned mostly on ordinary days. The variant, applied honestly, drags a whole card's deltas in one direction. Saturday's turf figures — which the chain leaves unadjusted by standing convention — leaned the other way, +0.57 on average. Same horses, same week: the sign of the average miss follows the adjustment rule, not the animals.

And then there are the misses no construction explains. Sweet Azteca went into the Filly & Mare Sprint with a projection of +1.45 and earned –11.44, a delta of –12.89, finishing last of 11. The report's verdict on that line is short: a genuine flop, not an artifact. At the other end, Kathynmarissa won her race with an actual of +3.67 against a projection of –1.56. A figure is a summary of what a horse has been. The distribution of what a horse does next has a wide middle and real tails, and any use of a figure that forgets the tails will be reminded of them at the windows.

Is the variant worth its trouble?

Everything above might read as an argument for throwing the variant away and betting raw figures. The lab tested exactly that, head to head, under a rule written down before the results were read. The test deserves a careful telling, including its own flaw, because the flaw is part of the lesson.

The registered version of the test aimed at the lab's production line store — and collapsed on coverage. That store is a 2025–2026 product artifact; the test population was the lab's research corpus, and only a small share of its chart-confirmed runners had a scored store line before their race. Only a handful of eligible races remained. On those races the two arms never disagreed, and the test could distinguish nothing. The lab recorded that result, then ran the identical frozen protocol — same arms, same last-three-lines window, same ranking rule, same decision rule — on the corpus's own figure chain, the build that had produced the corpus's figures in the first place. The substitution was written up before the supplementary outcomes were computed. Any honest citation of this test carries the substitution with it.

Each race, rank the field twice — once by the median of each horse's raw figures from its last three lines, once by the median of the adjusted figures from the same lines — and bet the top horse in each ranking.

The adjusted ranking won more races than the raw ranking. It placed the actual winner higher on average and put the winner in its top three more often. Head to head, in the races where the two rankings named different horses, the adjusted pick won more often than the raw pick, by a clear margin. On money, both books lost about a fifth of their stake; the paired difference favored the adjusted ranking, and its confidence interval included zero.

Read those results at the four-claim discipline this book keeps. The ranking claim is strong: three separate ordinal measures point the same way, and the duel margin is wide. The probability claim was not the test's subject. The wagering claim

is weak: the paired money edge was small, with an interval that cannot exclude no difference at all. And both books lost heavily, because a ranking of past-performance figures, however built, is still just a ranking — Chapter 1's crowd beats it, and the takeout beats everything that merely matches the crowd.

The cuts inside the test teach the practical version. Split the races by how large the variants in the horses' recent lines were. Where the variant hardly moved, the adjusted ranking still won slightly more often but returned less — the paired difference ran the other way: on quiet days, the adjustment is mostly reshuffling, and its extra winners paid shorter prices. In the middle range the paired difference favored the adjusted ranking, and where variants ran large it favored the adjusted ranking by more, with the duel net positive at every level. The variant earns its keep exactly where you would hope: on the days when the track was genuinely helping or hindering, which are the days a raw figure lies to you with a straight face.

The rebuilds that failed

A result like Sovereignty's invites a specific suspicion: maybe the variant is right in general but built wrong in the details — pooled across the whole card when it should be split, applied at full strength when it should be softened at the distances and in the company where elite horses run. The lab tested those rebuilds too, each under its own frozen rule, on the same research corpus. This section is short because the results were.

Split the variant by distance band — one variant for sprints, one for routes, instead of one pooled variant per surface. The challenger gained little on the paired comparison, lost the head-

to-head duel narrowly, and won slightly less often than production. Verdict: not supported. Its gains were real but narrow — routes, and high-variant days — and its losses in sprints and mid-variant races cancelled them inside the same population.

Shrink the variant toward each race's own sample — let a race's starters pull the day's variant partway toward what that race alone suggests. This arm posted the best placement numbers of any construction the lab tried: it put the winner highest on average and found the winner in its top three most often. It also returned the worst money of the three, because the winners it gained were shorter-priced than the winners it lost. A cleaner ranking and a worse book, in the same arm — the four claims separating again, in a single test. Verdict: not supported.

Damp the variant — apply it at three-quarters strength, half strength, and so on. Applied across the whole population, damping was not supported at any level; the winner's average placement got worse at every step down, monotonically, which is about as clean a signal as a test like this gives. One narrow cell did clear the frozen rule: races at nine furlongs and beyond, at three-quarter strength. The lab judged that cell too fragile to adopt, and it was not adopted. It stands as a recorded result, not a reason to soften the variant on your sheet. The Classic that prompted the question is not even in that test's population — the 23 Classic cards are historical charts outside the research corpus, so the test answers the general question, not Sovereignty's.

The production figure — raw minus the full pooled variant — stands against all three challengers. That is not a claim that it is the best figure that could be built. It is a claim that three specific, plausible improvements were given full-count tests and failed them. In this lab, that is what "stands" means.

The constant underneath: an audit

One more layer sits under the variant: the track constant itself, the "typical" each raw figure is measured against. In November 2026 the lab audited its production constants file — built from the 119,105-line store only — cell by cell. The audit's job was to find out whether the gaps a handicapper might blame for a strange figure were defects or just scarcity.

The store covers 1,089 combinations of track, distance, and surface. Of its scored lines, 87.8% are priced against their own exact cell. Most of the rest fall back to a broader rung of the ladder — the track's constant across distances, or the track-and-surface constant — because a lookup cell requires at least 30 lines, a standing minimum the lab applies so that a "typical" is never built from a handful of races. Among lines whose surface is classified, only 2.79% fall below their own cell. The large fallback headline a quick count produces comes almost entirely from lines whose surface was never resolved, which cannot form cells at all. Scarcity, in other words, wearing the costume of a coverage problem.

The audit's named case was Keeneland at ten furlongs on dirt — Sovereignty's exact cell. The store holds zero lines for it. Keeneland cards that race, in this era, essentially only when it hosts the Classic, and a race enters the store only through past-

performance sheets pulled after it runs. So the Classic prices at the next rung down: the Keeneland dirt constant, built from 942 lines. Is that rung misleading the figure? The audit checked it against true ten-furlong dirt cells at other tracks; Saratoga's sits within a hair of it, and the estimated error in the Classic's pricing works out to about 0.05 figure points — a twentieth of a point, in a table where variants swing three points either way. The audit's verdict was three words long: no defect found. The cell everyone suspected was the one part of Sovereignty's number that was nearly exactly right.

That is a pattern worth keeping beyond this chapter. When a figure looks wrong, the eye goes to the exotic explanation — the missing cell, the famous horse, the big race. The measured causes are usually the plain ones: a large variant doing its documented work, and a scale doing what scales do.

How to read a figure on Saturday

Everything in this chapter compresses into a short working discipline.

A figure is an ordering, not a price. The raw-versus-variant test is the cleanest statement of this in the lab's records: the better construction improved the ranking on every ordinal measure and still produced a win book that lost heavily, with a money edge whose interval includes zero. A figure can tell you which horse has been fastest. It cannot tell you whether the crowd has already charged you for that speed — and it usually has. Ranking, probability, and return are different claims; a figure speaks to the first and is silent on the other two.

Ask what the adjustment did before you argue with the total. When a figure surprises you — too low for a performance you watched, too high for one you dismissed — the useful question is not "is this number right?" It is "how much of this number is the horse, and how much is the day?" If the sheet or the figure-maker will show you the variant, look. A two-point gap between two horses means one thing when both earned it on ordinary days and another when one of them earned it on a day the chain charged three points to the track. Sovereignty's +0.81 and his +2.85 are both true statements. Betting as if only one of them exists is how figures mislead careful people.

Expect the middle of the distribution, respect the tails. The Breeders' Cup control check put numbers on this: the median starter ran within about half a point of its projection, and the card still contained a −12.89 and a +5.23. A figure is the center of a range of days the horse might have. The horse that "cannot lose on figures" is a horse whose range you have not imagined honestly.

Do not rebuild the number at the kitchen table. The lab, with full records and frozen tests, could not improve its figure by splitting the variant, shrinking it, or damping it. A player adjusting figures by feel — adding two points because the day "looked fast," forgiving a slow number because the company was strong — is applying a variant of one, estimated from memory, never tested. The published figure, with all its machinery exposed, is a better starting point than a private one with none.

And keep the Sovereignty line pinned above the desk, because it generalizes to every number in this book: **the argument about the figure was never about the horse.** It was about what a

figure is — a performance, minus a judgment about the day it happened on, expressed on a scale where zero means typical and the top of the historical range runs near three points. Read any figure that way — ask for its parts, name its universe, and grade its claims separately — and the column on the sheet becomes what it should have been all along: not an answer, but a well-built piece of evidence.

Chapter 7 — Pace is a shape, not a number

Every race tells two stories. One is about ability: who is fast, who is fit, who belongs in this company. Chapter 6 took that story apart. The other story is about geography: who stands where, at each point of the running, and what that costs or buys them. That story is pace.

Most players meet pace as numbers — an early figure, a late figure, columns to sort the way they sort speed figures. This chapter argues that pace numbers are the wrong shape for the thing they describe. Pace is not a property of a horse, the way a turn of foot is. It is a property of a *field on a day*: of who else entered, who else wants the lead, and how the early running reshapes the late running. Read it as a picture and it is honest and useful. Read it as a ranking and it will lie to you — and the lab's records show how, and in which direction.

Two numbers and a picture

A past-performance line gives each horse pace figures for its past races. In the sheets the lab works from, the early figure measures how fast the horse ran in the opening part of a race, and the late figure measures how fast it ran from the late calls to the wire. Each horse also carries a running style — the sheets print it as a letter pattern — and a set of early-speed points, zero to eight, summarizing how often and how sharply the horse has gone to the front early in its races.

None of those numbers means much alone. An early figure that towers over one field is ordinary in another. So the working move, in the lab's product and in any careful player's notebook, is to convert each horse's pace figures into its standing *inside this field*: a percentile. Your horse's early figure sits at the 80th percentile of today's field — faster early than four-fifths of its rivals, on paper. Its late figure sits at the 30th. Now you can draw the race: two bars per horse, one for early position, one for late, each showing where the horse stands relative to the field it actually faces.

That drawing is what the lab prints. Its pace graphs are a *projected-pace display*: take each horse's early and late standing, temper the picture with the market's own view of the horse (the crowd's ranking carries real information about who will actually be sent, and who will actually finish), and draw the shape the race is projected to take. Two cautions ride on that picture, and the lab prints its work under both of them.

First, it is a projection, not a prediction. It projects how the race may unfold if the horses run to their past patterns. It does not predict the result, and it was never built to. Jockeys change plans at the gate. Horses stumble, get bumped, get sent by a trainer whose instructions nobody published. The projection is a sketch of the likely geography, drawn before the gates open, by someone who knows the sketch is not the territory.

Second — and this is a lab rule with a section number, not a slogan — the picture is annotation. In the production selection protocol, the morning-line order is the ranking anchor, full stop. Pace graphs sit beside the ranking and explain it; they do not reorder it. A separate probability model the lab runs (a blend that prices filed picks) carries its own prohibition, written into

the protocol in so many words: the blend never ranks, never reorders, and never becomes the selection method. Chapter 6's lesson was that a speed figure orders a field without pricing it. This chapter's tests ask the harder question about pace: can the shape of a race reorder the field *better* than the market's order does? The answer, measured several different ways, is no — and the ways matter, because some of them fail by a little and some fail in the direction opposite to the story everyone tells.

The vocabulary of shape

Before the tests, learn the words, because the words are the part of this chapter you can use on Saturday whether or not you ever bet a pace angle. The lab's race-shape module sorts a field's projected early running into three classes, and names five recurring situations inside them. The definitions below are the lab's frozen ones, stated plainly. Note the ending before we start: the lab retired this type system as a set of betting flags after its first version — the types survive as vocabulary and as description, not as signals — and the chapter's tests will show why that retirement was the right call.

First, the three classes of early shape, read off the projected early order:

- **Soft.** At most one horse in the field is genuinely early-committed — built to run at the front — and the projected leader stands clearly clear of the second horse. The front is uncontested and the leader can expect to set its own tempo.
- **Hot.** Three or more horses want the same early ground, or three or more sit within striking distance of the leader.

The front is contested; the early tempo is likely to be fast and expensive for everyone who pays for it.

- **Honest.** Everything between: some company up front, no war.

Inside those classes, five named situations:

Solo lead. Exactly one early-committed horse in the field — a horse whose style is to run at the front, or a forwardly placed horse with strong early-speed points — and that horse is the projected leader, standing at least a clear gap ahead of the next horse in the projected order. The loneliest and, in bar-room handicapping, the most beloved shape in racing: one horse, alone, unhurried, with the race's tempo in its rider's hands.

Solo buried. The mirror image. The field's only one-dimensional early horse — a pure front-runner, nothing else in its repertoire — is the projected leader, but its *recent* speed figures fall well short of the field's competitive range while an older figure shows it once belonged. The lead is there for it. The horse that used to take the lead and stay may no longer be in it.

Soft front. The leader in a soft shape whose late figure is ordinary — at or below the middle of the field, roughly. It will get its trip. The question the shape asks is whether an ordinary finisher with a perfect trip beats better horses with worse ones.

Pocket trip. In a soft shape, or an honest one with a single clear leader, exactly one horse projects to sit in the pressing slot — right behind the leader, within about three lengths — and that horse's late figure beats the leader's. It gets the run of the race: close enough to strike, saving the ground and the effort the

leader spends, with the finishing kick to pass it. If you have ever watched a race and said "that horse got the trip," this is the trip, named in advance.

Meltdown. A hot shape, and a horse that runs from behind — a closer, or a presser-closer — holding the strongest late figure among the off-the-pace horses, with figures good enough to belong, and a history of running its race when the pace in front of it is fast. The leaders spend each other; this horse collects what they drop.

There is also a label for the common building block: a **dual early** horse, a forwardly placed runner with strong early-speed points — the kind that creates contested fronts when two or three of them land in the same race.

Read those five again and notice what they are. Every one is a *story with a mechanism*: a reason the race's geography might help one horse and tax another. That is exactly what makes them valuable as description and dangerous as evidence. A story with a mechanism feels like knowledge. The feeling is not the knowledge. The lab's whole pace program can be summarized as: keep the stories, test the claims, and publish which is which.

The comfort zone, and a warning label

Two working details from the frozen definitions show how seriously the module takes its own picture — and how openly it labels the picture's weak points.

The first is distance, measured in lengths. To place horses in the projected order, the module converts a pace-figure gap into lengths with a frozen operational constant: four pace points to a length. It then asks, for each horse, whether the spot the

projection assigns it resembles the spots it has actually occupied. A horse's typical early position — the median of its beaten lengths at the first call over its last three starts, zero if it led — is its comfort zone in the lab's plain sense of the term. A projected gap that lands a horse near that typical spot is a race it recognizes. A projected gap that strands it far from home — chasing a lead it never chases, or buried behind horses it usually sits in front of — is a trip it has not shown it can run. The meltdown definition uses this directly: its comfort gate prefers horses whose typical early deficit is moderate, treats a typical deficit of ten lengths or more as troublesome under a fast pace, and wants the projected trip inside about two lengths of the horse's usual one. None of this is mysticism. It is the observation, made mechanical, that horses run their races from the neighborhoods they know.

The second detail is a warning label the module prints next to its own flags. A horse's early-speed points rest on its past races, and some pasts are thin or mismatched: fewer than three career starts, or a horse whose recent races came mostly at sprint distances stepping into a route, or the reverse. The module marks those horses' points unreliable and reports its results with and without them. Sit with that choice for a moment, because it is the opposite of how pace is usually sold. The tool that draws the picture tells you, horse by horse, where the drawing is guesswork. A casual player can borrow that habit for free: when you sketch a race by hand, mark the horses whose projected spots rest on two starts and a hope. The shape of the race is only as solid as its least-known participant, and the honest sketch says which one that is.

The sketch, tested as a sketch

The first test is the fairest one, because it asks the picture only to do the job it claims to do. A pace projection says: here is roughly who will stand where early. So — does it?

The lab checked its own sketch against the official charts of 207 settled races, 1,501 starters, from twelve days of its daily cards. Two versions of the projected early order went in: an older blend that mixed the early figure with the early-speed points, and a simpler revision that ranks by the early-speed points alone. The revision won plainly. Its projected order correlated with the actual early order at 0.427 against the blend's 0.387, and its projected leader was the actual early leader in 41.2% of the matched races — 61 of 148 — against 32.4% (48/148) for the blend. Drop the figure, keep the points, get a better sketch. Simplicity won a test it was actually entered in.

Now grade that result honestly, under the four claims. As association, the sketch is real: a correlation of 0.427 over 207 races is not noise, and a projected leader who actually leads two times in five is information you did not have without it. As a prediction, it is modest: three times in five, somebody else leads. Anyone who has watched a speed horse break slowly or a longshot get sent by an aggressive rider knows those three races personally. The sketch describes the geography a field *tends* to produce. It does not schedule the race.

So pace passes its first test: it describes how a race may run, measurably better than chance, and a simpler construction describes it better than a cleverer one. Keep that sentence. The failures start when pace is asked the next question.

The big test: do pace interactions reorder anything?

Here is the strongest version of the pace claim, built the way a believer would build it. Not raw pace figures — those are Chapter 6's territory — but *interactions*: features that encode the shape itself. The lab's pace-interaction experiment, preregistered and run on its full research corpus, built six of them for every starter: congestion (how crowded the horse's early neighborhood is), lone speed (how alone it projects to be), style crossed with pressure (does this horse's way of running fit the pressure the field creates?), and pressure crossed with distance (does the pressure matter more or less at today's trip?).

Those features were then given every advantage. They were added to the lab's strongest existing model — a locked probability model that already used the market and the standard feature set — and the combination was judged on the held-out races the builders had not fit on. If the shape of a race carries information about *who wins* beyond everything else, this is where it would show.

It showed — a little. The combined model's probability error improved by a small amount. That improvement is real in the statistical sense: its confidence interval excludes zero, and it appears in both halves of the holdout. It is also small — a fraction of the gain the base feature set itself makes over the market alone — and it fell short of the materiality bar the lab had frozen before fitting. Probability was the kindest of the four claims. The others were unkind.

In the only currency the ranking program pays in — the top pick's win book — the pace features bought nothing at all. The model with interactions and the model without named the same

number of winners. The interaction model disagreed with the favorite in far more races and converted all that extra disagreement into slightly less returned on the same stake. And the interaction features on their own, with no market and no base features to lean on, were worse than the market alone by a wide margin — a losing book and a head-to-head duel against the standard model lost heavily. Whatever signal lives in the interactions is texture on top of real information. Standing alone, it is not information.

The frozen gates required all four conditions. Two failed — the size of the probability gain, and the paired money comparison. Verdict: **not supported.** Field-level pace-interaction features do not add ranking information at the standard the lab set before it looked. The file closes; that is lab language for "we do not get to keep trying versions of this until one passes."

Sit with the shape of that result, because it recurs everywhere in pace research. The interactions moved probability mass in the right direction *on average* — a genuine, measurable association — and their changes to the actual picks were a wash. A claim can be true at the level of association and empty at the level of ranking, and empty again at the level of returns. Three of the four claims failed or fell short while the weakest one passed. A report that quoted only the passing one would be telling the truth and misleading you completely.

Two more ways to ask, two more answers

The interaction test is a statistician's version of the pace claim. The lab also ran the handicapper's versions, in a six-method program that gave pace two of its methods and a priority combination. These ran on the newer research corpus, against the plain morning-line order as the baseline to beat.

Pace efficiency. Take the ratio of a horse's late pace standing to its early pace standing — how much of its speed it retains late. Engineering language for an old idea: the efficient horse wastes nothing. As a rule — rank the field by that ratio and bet its top horse — efficiency named winners far less often than the favorite did, returned less than the favorite did, and lost its head-to-head duel against the favorite by a wide margin. The paired difference ran against the rule, with an interval spanning zero. Not supported.

The diagnostic underneath is the interesting part, and it points the wrong way. Split the *favorites* into thirds by their own efficiency ratio. The favorites with the lowest retention won most often. The middle third won less often. The highest-retention favorites — the efficient ones, the ones the ratio blesses — won least often of the three. As a description of favorites, efficiency runs backwards in this population. Chapter 11 takes that inversion apart in detail; here it stands as the chapter's cleanest warning that a ratio can be arithmetically sound, mechanically plausible, and pointed in the wrong direction on the only races that count.

Competitive pressure. Count, for each horse, how many rivals have early pace near or above its own — how much company it will have doing what it does. This one describes favorites

honestly. Split favorites by how much pressure they face, at the tightest setting the lab used: favorites with little company won most often; the middle band won less often; favorites facing heavy pressure won least often. That is a real gradient, and a casual player can use it as description: a favorite with three rivals built like it is a shakier proposition than a favorite alone on the front.

But description is not a pick. The rule the lab built on pressure — find the horse the pressure favors, a stalker or closer positioned to benefit — returned less than the favorite did, won far less often, lost its duel to the favorite by a wide margin, and posted a paired difference against it. Not supported. And the priority experiment, the one the program named first — efficiency gated by pressure, the two methods combined — also posted a paired difference against it and lost its duel heavily. Not supported, and per the preregistration, nothing combines further. The pressure gauge tells you the favorite's day may be hard. It does not tell you who beats it. Those are different sentences, and the distance between them is this book's recurring subject.

One more screen deserves its line, because it is the story every pace player tells first. **Lone speed, weak late:** the favorite is the field's clear early leader and its late standing is poor — the classic false favorite, alone in front with nothing left for the stretch. The screen that flags exactly that favorite found only a small group of them in the evaluation races. They won *more* often than favorites generally — at a return in line with the favorite baseline. The screen built to find weak favorites found a small group of strong ones. The group was too small to build a contrary theory on, and the lab built none. But as a search for

vulnerability it failed in the instructive direction: the story was coherent, the mechanism was plausible, the count said the opposite.

Its companion screen — the closer who leaves too much to do, in a big field — separated on the primary corpus and then reversed its money separation on the control corpus. A screen that changes its answer when the population changes is a description of a population, not a tool. Recorded; not adopted.

What pace is for

Given all of that, a fair question: why does the lab still draw the graphs, still classify the shapes, still print the picture on every card? Because the tests above do not say pace is useless. They say, precisely, what it is useless *for* — and what remains is worth having.

Pace explains races. After the fact and before it, the shape is the best short account of *how* a result happened or may happen: the lone leader who got brave, the contested front that set the race up for the closer, the favorite who never got the trip its figures required. Explanation is not a consolation prize. It is how a player builds the judgment that no single factor carries.

Pace describes vulnerability without naming a winner. The pressure gradient is the honest example: favorites under heavy early pressure win materially less often, and a player who knows that knows something true about the race — something that can justify a pass, a smaller unit, or simple caution — without possessing a replacement pick. Passing is a decision with value;

Chapter 17 prices it. What the gradient cannot do is promote the stalker behind the favorite into a bet. The rule built on exactly that promotion lost.

Pace keeps you out of specific trouble. A horse whose every good figure came with one trip — alone in front, in the pocket — facing a field that denies it that trip is a horse whose figure means less than it prints. That is a negative use, a discount, and negative uses are where shape earns its keep: not to find the winner, but to mark the favorites whose evidence is thinner than it looks. The screens warn you how weak even that power is — the lone-speed screen's flags won *more* often — so the discount stays a judgment, stated as one, never a rule in costume.

And in the lab's own product, pace holds one small, tested seat at the table: the early-pace percentile is one of five factors the consensus scores inside the morning line's top four. Mid and late pace asked for seats on the same panel, were tried, and earned none; the wider pace display stays annotation. That is the measured size of pace's contribution to picking winners in this operation: a voice, inside a committee, inside the market's shortlist — after every attempt to give it the chair came back not supported.

So read the picture the way the lab prints it. Two bars, a shape, a story about geography. Believe the geography — it tests real. Doubt anyone, including this book, who tells you the geography names the winner. The market's order survived the blend of figures and points, the six interaction features, the efficiency ratio, the pressure rule, and the combination of the last two —

not because pace is nothing, but because *how a race runs* and *who wins it* are two questions, and only one of them is pace's to answer.

Chapter 8 — Class: real, and mostly already priced

Every casual player learns the class story early, usually from the player next to them, and it usually sounds like this.

That horse just ran against better. Look at the company it kept. It was beaten six lengths in a stakes race, and today it is in with claimers. These horses cannot run with it. It will gallop.

Sometimes that horse does gallop. Class is not a fairy tale. Horses really do face harder and easier company, and a horse that held its own against strong fields often does handle a weaker one. If class were nonsense, this would be a short chapter: we would test it, watch it fail, and move on.

That is not what happened. Class, measured carefully, turned out to be real — the clearest description of favorite strength in an entire six-method research program. And the pick rule built on it lost.

Hold both halves of that sentence. This chapter is about why both halves can be true at once, and it is the cleanest example in this book of the discipline we set out in Chapter 1: association, ranking, probability, and returns are four different claims. A measurement can win the first claim outright and fail the other three. Class did exactly that.

What “class” means when nobody is selling you anything

Strip the word down. A race has a level. At the top sit stakes races, where owners pay to enter good horses and the purses are large. Below them sit allowance races, for horses that have not yet won much or that meet set conditions. Below those sit claiming races, where every horse in the field can be bought — claimed — for the price tag in the conditions. A horse entered for a claiming price is, in a plain sense, offered for sale at that price, and owners do not offer a horse for less than they think it is worth if they can help it.

So the level of a race tells you something about the horses the track expected to fill it. A horse dropping from allowance company into a claiming race is, on paper, meeting easier rivals. A horse rising from claiming company into an allowance race is meeting harder ones. Players call the first move a drop and the second a rise, and whole handicapping conversations consist of arguing about whether the drop is genuine (the horse is finding its level) or a warning sign (the connections are giving up on it and would be glad to lose it to a claim).

Notice what class is not. Class is not a speed figure. Two horses can run the same time at the same track, one against stakes horses and one against cheap claimers, and the identical number does not mean the two runs were equally hard to produce. The stakes horse was pressed by better rivals, set or chased a stronger pace, and fought better horses in the stretch. The claimer’s number came against company that folded earlier. Raw figures treat the two runs as equals. Every experienced

player knows they are not equals. That gap — between what the figure says and what the company implies — is the space class handicapping lives in.

The question for a lab is whether that gap can be measured instead of argued about. It can. The tool is older than computers, and it is called a par.

Pars: the typical number for the race

A par is a simple idea. For a given kind of race — this track, this distance, this surface, this class level — what figure does a typical starter run?

Say, purely for illustration, that the typical starter in a certain class of sprint at a certain track earns a figure of 80. That 80 is the par for that race type. It is not a winning figure and not a good figure. It is the middle of the crowd for that specific job. A horse that runs 84 in that race ran four points above what the race usually produces. A horse that runs 76 ran four points below it.

Now the two identical raw figures from the last section stop looking identical. Suppose the stakes race has a par of 92 and the claiming race a par of 78. The horse that ran an 84 against stakes company ran eight points *below* its race's par. The horse that ran an 84 against claimers ran six points *above* its race's par. Same number on the page, opposite stories once you ask what the race itself usually demands.

That adjustment — figure minus par — is the whole of the method tested in this chapter. The lab calls it class-adjusted performance, CAP for short. A horse's CAP asks: in its recent races, how far above or below the typical level of those races did this horse actually run?

One detail matters before we go further, because casual players meet a sloppy version of this idea everywhere. A par is only honest if it is built without peeking. If you compute the par for a race using the results of races run later, you have let the future leak into the past, and any test built on it flatters itself. The lab built its pars the slow way, and the next section describes the build because the build is part of the claim.

How the lab built the pars

The pars in this test come from the research corpus, with the larger corpus as the control. Every verdict in this chapter carries that scope.

The pars themselves were built from the corpus records in date order, expanding as the year went on. The par for a race was computed only from races run before that race's date. A race in June was judged against pars built from January through May. A race on the first date in the corpus had no earlier history to draw on; a small number of records races received no par at any level, and the test says so instead of quietly filling them in.

For each race, the par was the median figure of the starters in a comparison pool — median, not average, so one freakish performance does not drag the "typical" number toward itself. The lab looked for the pool in steps, from the most specific to the broadest:

- First, the full cell: same track, same distance band, same surface, same class group. That level supplied the par for most races.
- If that pool held too few starter values, the class group was dropped: same track, distance band, and surface.
- Then track and surface only.
- Then surface alone.
- If no level reached the value floor, the race got no par.

The distance bands were three: sprint (under 7 furlongs), mile (7 to 8.49 furlongs), and route (8.5 furlongs and up). The class groups were five, read off the race type: maiden, allowance (including starter and optional claiming), stakes (including handicaps), claiming, and everything else. The exact groupings were fixed before any outcome was read. That sentence appears in this book often, and it matters every time: nobody looked at the results and then chose the grouping that made class look best.

With pars in place, each horse's CAP was built line by line. Take one of the horse's recent races. Take the figure it earned there. Subtract the par of *that* race — the race the figure was earned in, not today's race. A horse that ran above the level of its own race gets a positive number for that line; a horse that ran below it gets a negative one. The lab used the horse's last three comparable lines. Comparable had a definition too: the same

surface family as today's race, and a distance within one furlong of today's. One usable line gave that line's value; two lines were averaged; three took the median, the middle value.

Not every horse has a CAP. A horse needs recent comparable lines, and those lines need pars. On the primary corpus, a clear majority of starters carried a CAP, and most favorites carried one. Keep that scope in mind. The diagnostic in the next section speaks about the favorites who carried a CAP, not about all favorites, and an honest chapter does not blur the two.

The diagnostic: class adjustment describes favorites better than anything else we tested

Here is the test shape, and it is deliberately modest. Forget picking winners for a moment. Take only the favorites — the morning line's top choice, the crowd's horse. Sort those favorites into three groups by their CAP: the bottom third, the middle third, and the top third. Then ask one question: how often did each group win?

This is an association question. It does not ask CAP to pick anything. It asks whether CAP knows something about the favorite you are looking at.

On the primary corpus, the answer was the steepest gradient in the whole six-method program. Favorites whose recent figures were earned well above the pars of their races won most often. Favorites in the middle group won less often. Favorites whose recent figures sat at the low end, relative to the races that produced them, won least often. The gap between the top and

bottom groups was material, across a large count of favorites. That is not noise-sized. That is a measurement separating strong favorites from soft ones.

The control agreed. On the larger control corpus, the same three groups won in the same order — lowest for the bottom group, higher in the middle, highest at the top — with a wider gap in the same direction. When a pattern shows up in the primary population and again in the larger control, the lab takes the pattern seriously. Class adjustment describes something real about favorites. Of the six methods in the program — pace efficiency, pressure exposure, consistency, trend, class adjustment, and separation — this diagnostic separated favorites the most steeply of all.

Now read the returns alongside the win rates, because it is where this book's discipline earns its keep. Every group lost money. The top-CAP favorites, the strongest group, lost at roughly the usual favorite rate. Even the best-described favorites in the program paid back roughly eighty cents on the dollar, because a favorite that wins more often at favorite prices still does not win often enough to cover what those prices pay. The diagnostic improved your *knowledge* of the favorite. It did not improve the *bet* on the favorite. Those are different claims, and the test shows both at once.

There is a second caution inside the same result, and a careful reader should catch it. The win rates rise in a clean staircase from bottom group to top group — but the returns do not. The middle group lost the least, not the top group. If CAP thirds were a betting system in disguise, the returns would climb with the win rates. They do not, because price sits between them: the

market is not blind to class, and stronger-looking favorites tend to be priced shorter. The staircase in win rate is real. The staircase you could take to the window is not there.

So at this point in the test, class has earned exactly one claim, and it is worth stating in the book's four-claim language:

- **Association: yes.** CAP is associated with favorite strength, steeply, on both populations the lab keeps.
- **Ranking: not yet asked.** Nothing so far says CAP can order a whole field better than the morning line does.
- **Probability: not tested.** The lab did not build a CAP probability model or grade its calibration. A win-rate staircase across three groups is not a probability claim, and this chapter does not promote it into one.
- **Returns: no, for the favorite groups themselves.** Every CAP third of favorites lost money at roughly the usual rate.

Most handicapping writing stops after the first bullet. The staircase looks like a system. The next step — the step that turns a description into a pick — is where this method, like every method in the program before it, met the market.

The pick rule: take the highest-CAP horse in the race

The rule was written down before outcomes were read, in the program's preregistration. It is the obvious rule, the one a reader of the last section is already forming: in every race, pick the starter with the highest CAP. If no starter in the race carries

a CAP, fall back to the morning-line favorite — that fallback fired in a minority of races. Break any ties by the morning-line order. Bet $2 to win. Count every race.

The comparison is the same one this book always uses. ML-BASE — the plain morning-line order, favorite every race — is the baseline. To earn a verdict of SUPPORTED, the new rule had to clear three gates, all fixed in advance: its paired difference against the favorite, race by race, had to have a confidence interval sitting entirely above zero; it had to beat the favorite in the head-to-head duel — in races where the rule and the favorite named different horses, the rule's horses had to win more of those races than the favorite did; and its return had to sit above the favorite's return.

The highest-CAP rule, R-CAP, cleared none of them. Its win rate sat well below the favorite's. Its return sat below the favorite's. Its paired difference ran against it, and the interval crossed zero. Its duel against the favorite was lost by a wide margin.

Verdict: R-CAP NOT SUPPORTED.

The control returned the same verdict. On the larger control corpus, the paired difference also ran against the rule, the interval crossed zero, the duel was lost heavily, and the return sat below the favorite's. No cross-corpus flip, no rescue in the larger population. The rule fails on both populations the lab keeps.

Look at the shape of the failure, because it repeats across this book and learning to recognize it will save you more money than any single method in these pages. R-CAP did not pick randomly. Its win rate was far above what a dart throws at a full

field. The rule found real horses — horses whose figures were honestly earned above the level of their recent races. But the duel tells the story the win rate hides. In the races where R-CAP and the favorite disagreed, the favorite won nearly twice as often as the rule's horse did. Head to head, on the exact races where class adjustment said "not that one, this one," it was wrong nearly twice as often as it was right. And the prices of the horses it did land did not pay for the misses: its return sat below the favorite's return.

Why? The chapter's title gives the answer. Class is real, and the market knows it. The morning-line maker knows which horse is dropping out of allowance company. The betting crowd knows it too, and so does every other player whose money sets the closing price. A class edge that everyone can read off the same past performances is not an edge by post time; it is an input. It is already in the price. What is left for the highest-CAP rule is the residue: the races where the highest-CAP horse is *not* the favorite, which are disproportionately the races where the market looked at the same class story and found a reason — a layoff, a surface question, a pace problem, a trainer pattern, something — to prefer another horse. The rule systematically overruled the market exactly where the market had extra information. It lost like a rule that does that loses.

This is the same shape as Chapter 1's lesson, one level down. The crowd's ranking is the pooled judgment of everyone, and class is one of the things that judgment is made of. Beating that ranking requires information the ranking has not absorbed. CAP, for all its descriptive power, is built from the same public

figures and the same public race conditions the ranking was built from. It reorganizes public information. It does not add to it.

What a casual player should actually do with class

A failed pick rule is not a useless measurement. Here is what CAP's record honestly supports, stated as use, not as a system.

Use class adjustment to grade the favorite, not to replace it. The diagnostic is the finding. A favorite whose recent figures were earned above the pars of its races is a stronger favorite than the raw win rate of favorites suggests. A favorite at the bottom of the CAP scale is a softer favorite. That knowledge can shape a pass decision, a decision about how much company a favorite needs in a multi-race bet, or a decision to look harder at the rest of the field. What it cannot do, on this evidence, is hand you the horse that beats the soft favorite. The test tried. The horse it handed over lost the duel by a wide margin.

Do not confuse a class drop with a class edge. The drop is a fact about today's entries. The edge would be a fact about today's prices — that this horse's chance is better than its odds say. The lab's result says the first fact is usually already inside the second. When you see the obvious dropper at a short price, you are not looking at a secret. You are looking at the most visible horse in the race, priced like it.

Ask what the par of the race was, not just what the figure was. This habit costs nothing and needs no lab. When a horse's last race shows a big figure, ask what kind of race produced it. When it shows a small one, ask the same question. A modest figure earned against a strong field and a big figure earned

against a weak one are closer than the page makes them look, and sometimes the order between them flips. You will not beat the market with that question alone — this chapter is the evidence — but you will stop being fooled by the raw number, and that is worth real money over a season of small mistakes avoided.

Let a strong diagnostic stay a diagnostic. The temptation after seeing a clean win-rate staircase is to build on it: combine it with pace, add a price filter, keep the top CAP horse only when the favorite's CAP is low. The preregistration for this program allowed combining only after a method was individually supported, and R-CAP was not. That rule is not lab bureaucracy. Every added condition is another chance to fit the accidents of the races already counted, and Chapter 4 explains what that costs. The honest sentence is the short one: class adjustment describes favorites truthfully and does not pick winners better than the morning line. Both halves stay in the sentence.

The three lessons of class

First: class is real. Measured against pars built honestly from prior races only, horses whose figures were earned above the level of their company win more often. The favorite version of that statement is the steepest diagnostic gradient in the lab's six-method program, and it held on both populations the lab keeps. Do not let anyone tell you class is a myth old players hide behind.

Second: real is not the same as available. A truth the whole market can see is a truth the price already contains. The pick rule built on the program's best diagnostic lost to the morning-

line order on every gate — paired difference, duel, and return — on both populations the lab keeps. R-CAP is NOT SUPPORTED, and the size of the diagnostic staircase does not soften that verdict. It sharpens it. If even this measurement cannot beat the order, the order is absorbing a great deal of handicapping knowledge before you ever see a price.

Third: grade every idea on the claim it actually earned. CAP earned an association claim. It did not earn a ranking claim, it was never given a probability test, and its favorite groups earned no returns claim. The next time a method is presented to you with a beautiful sorting table — winners sorted by this, favorites sorted by that — ask which of the four claims the table proves. Then ask to see the other three. Chapters 6 and 7 asked that of speed and pace. This chapter asked it of class, the strongest describer we had. The next chapter asks it of two ideas that feel even more personal: the horse that always runs its race, and the horse that is getting better.

Chapter 9 — Consistency: the dependable horse and the improving horse

Ask a room of casual players to describe the horse they trust, and you will hear two different horses described as if they were the same animal.

The first horse always runs its race. Its past performances read like a metronome: figure, figure, figure, all within a few points of each other. It may not be the fastest horse in the field, but you know what you are buying. No surprises. Players call this horse dependable, honest, consistent — and they bet it the way you buy a tool that has never failed you.

The second horse is going somewhere. Its early lines are ordinary, and then the numbers start to climb: a better figure, then a better one, then a race that suggests the horse has figured the game out. It is lightly raced, or newly in form, or finally at its right distance. Players call this horse improving, and they bet it the way you buy a stock on the way up.

These are not the same bet. The first is a bet on low variation — the past was steady, so the future will be steady. The second is a bet on direction — the past was rising, so the future will be higher. Handicapping conversation blurs them into one warm word, *form*, and the blur hides a problem: a test that mixes the two questions can pass one and fail the other without anyone noticing.

The lab kept them apart. Method 3 of the six-method program asked the two questions separately, with separate measurements, separate diagnostics, and separate pick rules. One of the two measurements turned out to be weak tea that did not even reproduce. The other produced a clean, honest diagnostic — and then the worst pick rule in the entire program.

This chapter walks through both, in the order the lab met them. As in Chapter 8, keep the book's discipline in front of you the whole way: association, ranking, probability, and returns are four different claims. A horse can be genuinely improving — that is a description — while "bet the most improving horse" loses money faster than any other rule we tested. Description is not selection. That sentence ran Chapter 8. It runs this one twice.

The measurements: spread and slope

Both measurements start from the same raw material: a horse's recent speed figures, but only the comparable ones. Comparable had a fixed definition, set before outcomes were read. A past line counted as comparable if it was run on the same surface family as today's race — dirt with dirt, turf with turf, synthetic with synthetic — and at a distance within one furlong of today's. A turf sprint three months ago says something about today's turf sprint. It says much less about today's dirt route, and the test refused to pretend otherwise.

Each measurement needs five comparable lines. That requirement has a cost, and the chapter states it up front: on the primary corpus, about half of all starters carried at least five

comparable lines. The other half of all starters are silent in this chapter. Among favorites, only about half carried each measurement. Young horses, first-time starters, horses switching surface or distance — the very horses casual players argue about most — are largely outside these tests, because they do not have five comparable lines to measure. A finding about measured horses is a finding about measured horses. It does not extend itself to the unmeasured half of the game.

Consistency, measured as CV. Take the horse's last five comparable figures. CV — the coefficient of variation — is the spread of those five figures divided by their average. A horse whose figures cluster tightly around its own average gets a small CV. A horse whose figures swing — a big number, then a dud, then a middling one — gets a large CV. Dividing by the horse's own average matters: a fast horse and a slow horse can be equally steady, and CV grades the steadiness, not the speed.

Improvement, measured as Trend. Take the same five comparable figures, in date order. Trend is the average of the two most recent figures minus the average of the fourth- and fifth-most recent. A positive Trend means the horse's newest work sits above its older work. A negative Trend means it has been sliding. The middle line of the five is left out of both averages, so the comparison is between the horse's recent self and its earlier self, with a small gap between them.

Notice that neither measurement looks at class, pace, or price. That was deliberate. The program tested each method on its own information before anything was combined, and combining was allowed only after a method earned support on its own. Neither of these did, so nothing in this chapter was ever

combined with anything. What you are reading is each idea standing alone in front of the same judge every idea in this book faces: the morning-line favorite.

The dependable horse: a shallow slope that vanishes

Start with the diagnostic, as Chapter 8 did. Take only the favorites that carry a CV. Sort them into three groups by CV: the most consistent third, the middle third, and the most volatile third. Ask how often each group won.

The direction is the one the dependable-horse story predicts. The steadiest favorites won a little more often than the wildest ones. But look at the size of it: the whole gradient, across a large count of favorites, was small. Compare Chapter 8's class staircase, which spanned a much wider gap on the same population. This is a slope you could walk down without noticing your feet tilt.

And then the control took even that away. On the larger control corpus, the three groups won at nearly the same rate — the middle group highest, no staircase at all. The lab's report states it plainly: the CV gradient is strictly monotone on the primary corpus but shallow, and it does not reproduce on the control. In the book's language, that is an association claim that failed its replication check inside the same study. Whatever steadiness does for a favorite on the primary population, the larger count does not confirm that it does anything at all.

Grade it on the four claims before moving on:

- **Association: weak on the primary corpus, absent on the control.** The honest summary is "not established." A

pattern that appears in one population and dissolves in the control has not earned the word *finding*.

- **Ranking, probability, returns: not earned by the diagnostic.** No probability model was built or graded, and every CV third of favorites lost money, as favorite groups do.

The preregistration had a fixed rule for what happened next: a ranking rule runs only if its diagnostic separates — defined in advance as a strictly monotone win-rate staircase across the thirds on the primary corpus, with enough favorites carrying the measurement. CV met that letter of the test on the primary corpus, so the rule was run. The rule, R-CV, is the one the story suggests: in every race, pick the starter with the lowest CV, the most dependable horse in the field, with ties and empty fields handled by the morning-line order. In a substantial minority of races no starter carried the five comparable lines the measurement needs, and the rule fell back to the favorite there.

Against the same baseline used throughout this book — the morning-line favorite — R-CV won far less often than the favorite did, returned less than the favorite did, posted a paired difference against it with an interval spanning zero, and lost its duel by a wide margin.

Verdict: R-CV NOT SUPPORTED. The control agreed: on the larger control corpus, the paired difference also ran against the rule, the interval spanned zero, the duel was lost heavily, and the return sat below the favorite's.

The duel deserves a sentence of its own, because it is the part of the record a win rate can hide. R-CV's win rate sounds merely modest. But in the races where R-CV and the favorite named

different horses, the favorite won far more often than the rule's horse did. The most dependable horse in the field, chosen race after race, lost its head-to-head argument with the favorite by a wide margin. Dependability, it turns out, is a property the market prices without difficulty. The steadiest horse in a race is usually steady at a level the rest of the field can beat, and the crowd's order already knows which level that is.

Why does the dependable horse disappoint as a pick? The measurement itself contains the trap, and it is worth seeing because casual players walk into it weekly. CV grades steadiness relative to the horse's own average. It says nothing about whether that average is good enough today. A horse that reliably runs a figure too slow for this field has a beautiful CV. It will reliably lose. Consistency is a statement about the *width* of a horse's likely performance. A bet needs a statement about its *height* — and about its price. R-CV selected on width alone, and width alone is not a ranking.

The improving horse: a true diagnostic and the program's worst rule

Trend is the stronger half of this chapter, and the more dangerous one, because its diagnostic genuinely works.

Same test shape. Take the favorites carrying a Trend. Sort them into thirds by Trend, from the most negative (sliding) to the most positive (improving). Count the wins.

The staircase is clean: the sliding third won least often, the middle third won more often, and the improving third won most often. A favorite whose recent figures stand above its older ones wins materially more often than a favorite moving the

other way. And this time the control confirmed it. On the larger control corpus, the thirds won in the same order — sliding lowest, improving highest — at nearly the same size. Of the diagnostics in this chapter, Trend is a real association. Rising favorites are stronger favorites. Sliding favorites are weaker ones. That description is true on both populations the lab keeps, and this chapter does not take it back in what follows.

What follows is about a different claim.

The contingent rule, R-TREND, was written the obvious way: in every race, pick the starter with the highest Trend — the most improving horse in the field. Ties and empty fields fell back to the morning-line order; in the same substantial minority of races where no starter carried the five comparable lines, the rule took the favorite there, exactly as R-CV did.

Against the favorite, R-TREND won far less often, returned less — the worst return of any rule in the program — posted a paired difference against the favorite, and lost its duel by a wide margin.

Verdict: R-TREND NOT SUPPORTED. And not narrowly. Its paired confidence interval sits entirely below zero. Most failed rules in this book fail with an interval that straddles zero, leaving room to say the evidence was merely inconclusive. R-TREND leaves no such room on either population the lab keeps: on the control corpus the paired interval also sits entirely below zero, the duel was lost heavily, and the return sat well below the favorite's. R-TREND is the worst ranking rule in the six-method program — the only one whose paired interval sits entirely on the losing side on the primary corpus.

Sit with the contrast, because it is the central teaching of these two chapters in its sharpest form. The Trend *diagnostic* separates favorites on both populations. The Trend *rule* loses to the favorite by more, and more certainly, than any other rule the program tested. Both statements come from the same races, the same measurement, the same run.

How can both be true? The diagnostic and the rule ask different questions about different horses. The diagnostic looks at the favorite — a horse the market has already judged to be the best in this field, at this distance, on this surface, today — and asks whether its recent direction adds information about *it*. It does. The rule ignores the market's judgment entirely and asks a whole field of horses a single question — who has risen the most? — then bets the answer. The answer is usually a different horse from the favorite: in most races, R-TREND's pick was not the favorite. And there is a reason the most-improved horse in a field is so often not the best horse in it. Improvement is measured from a horse's own baseline. The horse with the largest rise is frequently a horse whose older lines were poor — cheap races, wrong distances, troubled trips — and whose recent lines merely bring it up toward the level the favorite already occupies every week. Rising the most and standing the highest are different properties. The favorite is usually the horse standing highest. The rule kept betting the climb and ignoring the altitude.

The duel makes the cost concrete. In the disagreement races, the favorite won far more often than the improving horse the rule substituted for it — nearly three favorites won for every improver. And the improvers that did win did not pay enough to

rescue the arithmetic: R-TREND's return was the worst of any rule in this chapter, worse than R-CV's and worse than the class rule's in Chapter 8, against the same favorite baseline.

Grade Trend on the four claims, carefully, because the grades differ and the difference is the lesson:

- **Association: yes.** Rising favorites win more often than sliding favorites, on the primary corpus and on the control. This is a description of favorites, and it is true.
- **Ranking: no.** Ordering a field by Trend and taking the top horse loses to the morning-line order, decisively, on both populations the lab keeps.
- **Probability: not tested.** The lab built no Trend probability model and graded no calibration. A three-step win-rate staircase across favorite thirds is not a probability for any individual horse, and this chapter does not use it as one.
- **Returns: no.** The rule's return is the program's worst, and even the favorite groups in the diagnostic — including the improving third — lost money.

A handicapping book written the usual way would print the diagnostic, call Trend a factor, and move on, leaving you to discover the rule's record with your own money. The lab's record says the factor and the rule are different objects. One is true. The other is expensive.

The peak that was not a fluke: the V3 screen

One more test belongs in this chapter, because it is the consistency story told from the other side — and because it failed in the most instructive direction a test can fail: backwards.

Players distrust a spike. A horse runs the best figure of its life, far above anything it has done before, and the talk at the track writes the race off. It got an easy lead. The track was playing fast. It will bounce — that is the word players use — back to its ordinary level, and anyone who bets it off that one race is chasing a fluke. The lab built that story into a screen, V3, as part of the same Phase 2 run. A screen is not a pick rule. It does not choose a horse. It flags a group of favorites and asks whether the flagged group is weaker than the rest — a vulnerability test, judged on the flagged group's win rate and return against the unflagged group, with counts.

V3 flagged the favorite when its last five comparable figures contained a peak far above its own middle: specifically, when the best of the five minus the median of the five was 20 points or more. That threshold was not chosen by looking at outcomes. It was fixed from the pooled distribution of the measurement itself, before results were read — across all favorites, the typical peak-to-median gap was well below that mark, most favorites sat well below it, and about nine in ten sat at or below roughly that level. Twenty points marked the extreme tail: a peak that towered over the horse's own range.

On the primary corpus, V3 flagged a modest group of favorites. If the bounce story were true of this population, those favorites should have won less often than the unflagged favorites, and

lost more money. They did the opposite: the flagged favorites won *more* often than the unflagged ones, and lost *less* money. The screen points the wrong way. The report's verdict is exact: on the primary corpus the flagged group outperformed on both measures, the same inversion an earlier screen in the program produced, and on the control corpus the flagged and unflagged groups were indistinguishable. **V3 does not identify vulnerable favorites.**

Read that result with care, because it is easy to over-claim in both directions. V3 does not prove that a peak is a good sign; on the control corpus it proved nothing at all, and a screen carries no adoption either way. What V3 does is remove a story. On the primary corpus, the favorites whose last five figures contained a towering peak were, as a group, slightly stronger than average — not weaker. One plausible reading is that a horse capable of producing a figure far above its own median has demonstrated a ceiling that its rivals may not have, and the market, which can see the same peak, has already weighed the bounce story in the price. That reading is an interpretation, offered as one, not a lab finding. The lab finding is narrower and, for a player, more useful: do not cross out a favorite because its best recent race looks too good. On a full count of races, that cross-out removes winners at least as often as it removes losers.

There is a habit hiding inside the bounce story that this screen exposes, and it reaches beyond this one test. The story feels like prudence. It presents itself as the disciplined move — distrust the outlier, demand the horse do it again, bet the dependable profile instead. But notice what the first half of this chapter found about the dependable profile: CV's gradient was shallow and did not reproduce, and the dependable-horse rule lost its

duel by a wide margin. Put the two halves together and a pattern appears. In this population, the market has already done the obvious discounting. The spike is discounted. The steady horse is discounted. What remains after the discounting is close to a fair price on both, and a rule that systematically fades the spike or systematically backs the metronome is not exploiting an error. It is paying the takeout on a preference.

What a casual player should actually do with form

Use Trend the way the evidence allows: as a note on the favorite. An improving favorite is, on both populations the lab keeps, a somewhat stronger favorite than a sliding one. If you already liked the favorite, its rising figures are a small confirming fact. If the favorite's figures are sliding, that is a small reason to look at the race again. It is not a reason to bet the field's biggest riser instead. The test ran that exact substitution across most of its races, and it was the worst trade in the program.

Stop grading horses on steadiness alone. A consistent record tells you the range of what a horse does, not whether the top of that range wins today's race. Before you admire a metronome past-performance page, ask the only question consistency cannot answer: is this horse's *usual* figure good enough against *this* field? If the answer is no, the horse will lose dependably, and its low CV will be no comfort at the window.

Do not fade a horse for one big race. The V3 screen flagged favorites on the primary corpus for exactly that sin, and the flagged group won more often and lost less than everyone else.

One big figure is information — about the horse's ceiling, at least — and the price already reflects the suspicion you were about to apply for free.

Notice how little of the field these ideas even cover. About half the starters on the primary corpus had no CV and no Trend at all, because five comparable lines did not exist for them. Any approach built on form cycles is an approach that is silent about half the horses you will ever look at — every young horse, every surface-switcher, every shipper at a new distance. A method that cannot speak about half the field cannot be the spine of a handicapping routine. It can only ever be a note in the margin, and this chapter's results say the margin is where it belongs.

The three lessons of consistency and improvement

First: the two horses are different horses, and they earned different verdicts. Consistency, measured as CV, produced a shallow gradient on the primary corpus that the control did not reproduce at all. Improvement, measured as Trend, produced a clean staircase on both populations the lab keeps. Lumping them together as "form" would have hidden both facts: the weak one's weakness and the strong one's strength.

Second: a true description can carry the worst rule in the program. R-TREND's paired interval sits entirely below zero on both populations the lab keeps. R-CV lost its duel by a wide margin, and R-TREND lost its duel by a wider one. Describing the favorite's direction never named the horse that beats the favorite, because the most-improved horse in a field is usually climbing from a lower floor than the favorite already stands on.

Altitude beats slope when the two disagree, and the morning-line order — which is, in large part, an altitude ranking — won both arguments head to head.

Third: beware the story that flatters your caution. The peak-outlier screen was built from the most reasonable-sounding caution in racing — that towering figure was a fluke — and on a full count it pointed backwards: the flagged favorites won more often and lost less than the unflagged field on the primary corpus, with no separation at all on the control. A story can sound like discipline and still be a way of deleting winners. Test the cautious stories with the same suspicion you bring to the exciting ones. The lab did, in this chapter, three times. Class in Chapter 8 described favorites best and still could not pick. Consistency could not even describe reliably. Improvement described honestly and picked worst of all. That is the honest state of the tools — and it sets up the question Part II ends with: if stacking measurements like these on top of each other is not the answer, what does all the stacking actually cost?

Chapter 10 — The cost of complexity

Every handicapper has had the same afternoon. Your pick runs second. You go back to the past performances and find the thing you missed. The winner had the best late figure at the distance. Or the best record on an off track. Or a trainer pattern you did not weigh enough. The fix feels obvious: add a check for that. Next time, you will catch it.

Do that for a season and you own a system with forty checks. Do it for a few years and you own a system nobody can describe in one sitting, including you. Each check arrived with a story. Each story was true, in the one race that taught it. The question this chapter asks is what the pile of checks costs you — not in time, although it costs time — but in truth. A complex method is harder to test, easier to excuse, and more likely to be remembering the past than reading the present. Our lab learned this by building one.

The specimen on the table

Start with the honest specimen: our own selection method, the one whose record this chapter discloses. It files under a written protocol, Selection Protocol v3.17. Read the protocol as a casual player would read any elaborate system in a book or a tout sheet, and count the moving parts.

An anchor comes first. The ranking starts from the morning line's order. The method does not throw that order away; it works inside it. A consensus step then scores the morning line's top four. Five factors vote: the morning line itself, a handicap weight, a speed figure, an early pace percentile, and a draw

index that measures how a horse's post has performed in its kind of race. Each factor ranks the full field and awards 5 points to its first horse, 2 to its second, and 1 to its third. A horse outside the top four can take points off the table but cannot win the vote. The method's own earlier pick also sits in the box as a column worth 5 points. The highest total inside the four becomes the selection.

Around that vote sit the overrides. A price cap replaces a non-favorite official pick priced over 7/2 with the morning-line favorite; exactly 7/2 stands. Two exclusion cells file no pick at all: a turf race that is not a maiden race, and a turf race with nine or more filed starters. A counter-move can, under its own conditions, replace a pick or add a separate leg. The steps run in a fixed order — exclusions first, then the choice between two filed scenarios, then the cap, then the consensus — because each step changes the input to the next one.

That is the machine. It is not a straw man built for this chapter. It is a real, frozen, numbered method, written down in full, with a version history that records what changed and why. If complexity has a defense, this method should be able to make it: every part arrived through a test or a deliberate decision, and the whole thing can be pointed to, line by line. So grade it the way this book grades everything else. What did all of that machinery buy?

What the machine bought

The record printed here is the lab's own settled product record: the filed cards of November 2–4, 2026 — 111 races. Read the caveat before the numbers: three days prove nothing either way. The sample is small, and this chapter uses it as the product's own filed record, not as proof about the method.

On that record, 95 races carried a filed selection. A scratched selection counts as a refund, so the book holds 90 bets. The selections won 25 of those races, 27.8%. A two-dollar win bet on every selection staked $180.00 and returned $101.20. The net return was –43.78%. The machine loses money on this record. This chapter does not soften it.

The comparison is against the favorite on the same races. In the 85 races where both the selection and the favorite started, the selection won 24 races, 28.2%, and returned –41.76%. The favorite won 25 races, 29.4%, and returned –38.88%. The paired difference is –2.88 points against the selection. The selection did NOT beat the favorite on this record. Say so plainly: on the lab's own settled product record for those three days, the favorite won one more race and lost less money.

Then apply the second test, the one that separates a pick rule from a wagering claim. The lab ran a separate economics test on an earlier, four-factor form of the same consensus idea — a different pick object from the production selection, and this chapter keeps the two apart. That test's headline paired difference favored the test object, but its confidence interval crossed zero. It also failed the concentration checks the lab sets for an economics verdict: a single price band carried most of the net, and one part of the calendar carried more than the whole

net, because the other part ran negative. The lab's ruling was ANNOTATION ONLY. In the language this book uses from Chapter 1 on: the method holds annotation-grade status. It is a pick rule with evidence behind it. It carries no proven wagering meaning.

Hold those two facts side by side, because the space between them is the cost of complexity made visible. A simple method that produced the same pair of results would invite a simple verdict: interesting, unproven, move on. A complex method invites something else — a search through its parts for the one that must be responsible, and a repair, and another part. That invitation is the trap.

Each factor is another ticket in the noise lottery

Chapter 4 made the multiple-testing point in full, and this chapter only borrows it: if you test enough ideas against the same races, some of them will look good by luck alone, and a rule tuned on the races that judged it has not really been tested. Every factor you add to a system buys another ticket in that lottery. The draw factor's own adoption record shows the size of the draw. When the lab tested candidate columns for the consensus vote, it ran 151 arms. Thirty-five passed the frozen gates. Thirty-three of the thirty-five contained the same one factor. The lab recorded that honestly in the protocol: one family, not thirty-five discoveries. The single-change test — add the draw factor alone, change nothing else — is what the adoption rested on. A reader who met only the headline "35 of 151 arms passed" would carry away a picture of broad support.

The truth was narrower and, for that reason, stronger. Complexity multiplies arms, arms multiply chances to pass, and only a discipline that counts the arms keeps the count honest.

The same lottery runs inside a single system even when nobody runs a formal sweep. A player with ten factors can explain any result ten ways. The speed factor was right but the pace factor outvoted it. The weight factor found the winner but the cap removed it. Each explanation sounds like analysis. Most of the time it is the lottery paying out in narrative instead of numbers: with enough factors, one of them pointed at the winner in almost every race, and the system gets credit for a hit it never actually selected. The only score that cannot be gamed that way is the score of the final pick, counted over every race, against the favorite on the same races. That is the score in the section above. Everything else is commentary.

Each override re-opens a closed question

An override is a rule that fires after another rule has spoken and changes its answer. Overrides feel like wisdom: the general rule, plus judgment for the special case. In a tested method, an override does something specific and dangerous. It re-opens a question that a test already closed, in exactly the races where the temptation to re-open it is strongest — the races where the first answer looks wrong.

Watch the cap do it. The consensus vote exists because, on a full count of races, its disagreements with the earlier pick booked better than the pick they displaced. The cap then says: except when the winner of that process is a non-favorite priced over 7/2. In those races the favorite goes back in. That may be the

right trade — the cap was adopted as a deliberate weighing of win rate against return, not as an accuracy claim, and the protocol says so. But notice what the cap costs the record. From the moment it exists, the method's results no longer answer one question. They answer a blend of two: how the vote does, and how the exceptions to the vote do. When the numbers disappoint, which part failed? When they please, which part earned it? Each override doubles the explaining work and halves the meaning of the total.

The counter-move and the exclusion cells carry the same cost in different forms. A counter leg is its own bet with its own record, and it must be kept in its own ledger or it quietly subsidizes — or quietly taxes — the headline line. The turf cells are cleaner, because they remove races instead of changing picks, but they still change the population every later number describes. A method that declines two kinds of turf races is not being measured on the same sport as a method that plays them. None of this makes overrides wrong. It makes them expensive in the currency this book cares about: every override is a second hypothesis riding on the first one's test, and it needs its own count or it is an untested opinion wearing a tested method's clothes.

The protocol's own history contains the sharpest version of the lesson. An earlier version added a sixth profile edge based on connections — the records of the trainer and jockey. It was adopted after its test passed marginally and after a materiality check written into its own pre-registration had failed; the adoption was a judgment call, recorded as one. Then a full historical recheck on a corrected corpus reversed the evidence: the variant with the edge scored below its own baseline. The

next version removed the edge in full. Read that sequence slowly. A component entered on a marginal pass and a failed rider. It lived in the filing rules. It left when a better count said its basis no longer existed. The system worked — the removal is the discipline functioning — but the episode is the cost in miniature. For as long as the edge sat in the protocol, every filing it touched carried an assumption the evidence could not support, and nobody could see that from the results, because its races were folded into the whole. Complexity hides its weakest part inside its average.

The whole becomes impossible to falsify

Add enough parts and a method stops being a claim. It becomes a committee that can always blame a member.

Ask a simple method a hard question: did you beat the favorite, on the same races, by enough that the interval clears zero? It answers yes or no. Ask a forty-part method the same question and you get a meeting. The answer depends on which parts fired, which overrides engaged, which cells were excluded, and which version of the method was in force for which third of the sample. Every version change restarts the prospective clock for the changed part while the unchanged parts borrow age from the old record. The method is never quite the thing that was tested, and never quite new enough to demand a fresh test. That is not a testing problem with a clever fix. It is what complexity is: a structure that generates explanations faster than races can arbitrate them.

Our lab's answer was procedural, and a casual player can copy it at kitchen-table scale. Freeze the method as a numbered version. Never edit a version quietly; a change ships as a new number with its reason attached. Keep the accounting single and plain: one official pick per race, a backup chain written in advance for scratches, and separate ledgers for anything that is not the official pick. Judge the whole only on its final selections, counted in full, against the favorite. Those rules did not make the method profitable. They made its lack of profit legible — which is the precondition for learning anything at all. A method you cannot falsify cannot teach you. It can only accompany you.

Four more parts that did not earn their place

The selection method is the specimen that survived its tests at annotation grade. In the same week the lab froze it, four other constructions — each one a plausible, well-motivated addition to the machinery — faced the same standard on the same research corpus. The standard was the frozen one used across the lab: a candidate must beat the simpler version without it on a paired return comparison, win the head-to-head duel where the two disagree, and hold at least the simpler version's win rate. Read what happened to each. The pattern is the chapter.

The rank blend. Take a horse's raw speed figure, its variant-adjusted figure, and the market's ranking, and blend the three ranks into one order. The idea respects all three sources at once; it sounds like balance. The morning-line form of the blend posted a small paired edge against plain morning-line order, with an interval across zero. It found the winner less often than the anchor did. Where the two rank-1 horses differed, the anchor won the duel. The small return edge came from prices,

not from finding winners: the blend won less often at longer prices. A stricter reading of the same idea — sort by raw rank first, then adjusted, then market — was the worst arm in its test: its paired difference ran against it, it lost its duel heavily, and it won far less often. And the ceiling version, given the closing odds no bettor can know before the race, still lost the duel to the plain morning-line order. Three ranks blended produced a ranking worse than one of its ingredients used alone. Verdict in the lab record: NOT SUPPORTED.

The figure blend. Instead of blending the final ranking, blend the figure itself: nudge or remap each past figure using its rank against the raw figure, the adjusted figure, and the market in its own race. This one needs care, because its record is mixed, and this book does not flatten mixed records. The quantile form, which rebuilds a figure from the blended rank, failed on both its market forms: one form posted a paired edge but lost its duel, and the morning-line form posted a paired difference against it and lost its duel. The nudge form with morning-line information — the only form a bettor could actually use before a race — was NOT SUPPORTED, even though its duel ran slightly in its favor. The nudge form built on closing odds, GN-POST, passed its frozen gates on the primary corpus and confirmed on the control population. Its disposition is stated in words: GN-POST was held, released and adopted 2026-11-05 with the Classic exhibit re-graded, as an ANNOTATION-GRADE adoption. Its intervals cross zero on both populations, its usable pre-race form failed, and the exhibit it was built to repair — one famous winner whose adjusted figure looked too low — it did not repair at all, for a structural reason the test report states plainly. A within-race rank cannot see a charge that falls on the whole race at once. Complexity that passes a

narrow gate still has to answer the adoption question: what can you do with it, before the race, that the simpler figure cannot do? Here the answer was nothing yet for the pre-race form.

The breakout exception. The production figure charges every race on a card a pooled variant — an average correction for how fast the track played. The exception says: when one race's own evidence diverges from the pool by enough, charge that race its own number instead. It is an override in pure form, aimed at one known sore spot. The lab tested thresholds of half a point, one point, and a point and a half, plus a version scaled to each race sample's own noise. All four failed. Their paired differences ran against them, with one threshold near even, and every arm lost its duel and its win-rate comparison. And the exception was not exceptional: even the tightest gate fired on a substantial share of eligible races, and the loosest gate fired on nearly half. A rule that rewrites a tenth to nearly half of all races is not an exception. It is a second variant system, noisier than the pooled one it corrects, because a single race's sample is a median over as few as four horses. Pooling exists to average that noise away. The test priced the trade, at every width, and the price was negative. NOT SUPPORTED at every threshold.

The variant arms. Two more constructions attacked the same pooled variant from the construction side. The first split the pool by distance band, so sprints and routes on the same card carried different corrections. The second shrank each race's own sample toward the day's pool instead of choosing between them. The band split posted a small paired edge with an interval across zero, lost its duel narrowly, and found winners slightly less often than the pooled figure: NOT SUPPORTED. The shrinkage arm posted a paired difference against it, lost its duel

narrowly, and returned less than the pooled figure: NOT SUPPORTED. The shrinkage arm carries the chapter's cleanest small lesson. It ranked winners *better* than the production figure — its average winner rank was better, and it placed winners in its top three more often. By the ranking claim, it improved. By the returns claim, it lost more, because the winners its pull toward a race's own sample promoted paid less than the winners it displaced. A better ranking and a worse bet, in one arm. Chapter 5 separated those claims; this arm is why the separation is not pedantry.

Step back from the four. None of them was foolish. Each had a mechanism, a motivating example, and a builder who could state exactly what it should fix. Each was tested under a rule frozen before its results were computed, on the same population, against the simpler construction it proposed to replace. None earned its place as a pre-race improvement. That is the base rate a casual player should expect for additions to a handicapping method: the plausible repair, tested in full, usually pays nothing, and the few that pass a first gate arrive with intervals that cross zero and a usable form that fails. A system that admits parts on plausibility will accumulate all four of these and call the accumulation experience.

The counter-discipline

The lesson is not "keep it simple" as a slogan. Simple methods lose money too; Chapter 1's favorite is the simplest method in the world, and it loses a fifth of every dollar. The lesson is a discipline for admitting parts, and it runs in one direction only.

Every component must earn its place against the simpler version without it, or come out. Not against a straw man. Not against the favorite alone, if the component lives inside a larger method. Against the same method minus that one part, on the same races, under a rule written before the results are read. The draw factor entered that way: one change, tested alone, adopted alone. The connections edge left that way, a version later, when its count reversed. The four constructions above never entered as pre-race improvements, because their counts said no. The discipline is symmetric, and the symmetry is the point. A part that cannot be removed by evidence was never really admitted by evidence either. It was admitted by enthusiasm, and enthusiasm is not a test.

For your own handicapping, the kitchen-table form of the discipline fits on an index card. Write the method down, all of it, before the meet starts. Number it, even if the number is just a date. When you want to add a check, write the check as a change, and write what would make you remove it. Keep one score: your actual first pick, every race you play it in, against the favorite in the same race. If a part has no score of its own, it has no vote. And when a season ends, hold the removal meeting before the addition meeting. Ask of each part, in order: if this left, what would the record lose? The parts with no answer are the cost of complexity, itemized. Remove them first, and let the method that remains — smaller, duller, and honest about its losing record — be the thing you actually test next.

That is the ground this book stands on for its final part. The myths come next, and they fall faster once you see the pattern: a good story, a number that sounds like engineering, and a count that was never run. Chapter 11 takes the cleanest of them.

Chapter 11 — The efficiency myth

Few ideas in handicapping sound as solid as efficiency.

A car that uses less fuel to travel the same distance is a better car. A runner who wastes less motion runs faster for the same effort. Engineers measure the ratio of what comes out to what goes in, and the higher the ratio, the better the machine. So when a player divides a horse's late pace figure by its early pace figure and calls the result an efficiency rating, the word does half the persuading before the numbers say anything at all. A horse with a high ratio keeps most of its early speed late. A horse with a low ratio spends its speed and fades. Surely — the story goes — the keeper is the better bet, and the fader is the horse to stand against.

Our lab tested that story twice, in two different forms, on two different bodies of evidence. The ranking rule failed both times. The diagnostic underneath it — the part that was supposed to tell you which favorites are fragile — pointed the other way. This chapter walks through both tests slowly, because the gap between how the idea sounds and what it counted is the whole lesson. A ratio is not a law. It is a claim, and claims get counted.

First, separate the claims

This book has separated four claims from Chapter 1 on, and pace efficiency touches all four, so name them before the evidence arrives.

A ratio can show an **association**: horses with high values of it win more or less often than horses with low values, in some population. A ratio can support a **ranking** claim: picking the highest-ratio horse in each race names winners better than some baseline. A ratio can support a **probability** claim: its values translate into accurate win chances. And a ratio can support a **returns** claim: betting by it loses less, or wins, against the prices on offer.

The efficiency story slides across all four without stopping. It starts with a physical intuition — retention sounds like fitness — and arrives at a betting instruction. The lab's tests grade the claims separately. The ranking rule gets a paired comparison against the morning-line favorite on the same races, a head-to-head duel where the two picks differ, and a return on two-dollar win bets. The favorite diagnostic gets a simpler read: split the favorites into three equal groups by their own ratio, low to high, and count how often each group wins. If high retention marked strong favorites, the high group should win most. Keep those two scoreboards apart as you read. A measurement can describe a population honestly and still fail as a way to pick horses. This one does something stranger. As a description, it runs backwards.

The corpus test: the ratio in its first form

The research corpus carries no second-call figure of the kind the published past performances print, so the lab tested the method first in the form the corpus allows. For each starter, the corpus form divides the horse's late pace figure by its early pace

figure: PEI_c = lpf / epf. A value near one means the horse's late figure nearly matches its early figure. A lower value means the late figure falls well short of the early one.

The rule, written down and frozen before any outcome was read, is the rule a believer would write. It is called R-PEI in the lab record. In each race, select the starter with the highest PEI_c. Break ties by the morning-line order. Bet it to win at two dollars. Compare everything against ML-BASE — the plain morning-line favorite in the same race — on the same races.

The population is the research corpus. Name the population whenever you cite the result, because the result belongs to it. One race in the records carries no starter-level morning line, so no baseline pick exists for it and no rule can be scored there.

The baseline first. The morning-line favorite in the evaluation races won about a third of the time and lost about a fifth of the money on win bets. That is the familiar Chapter 1 favorite: right about winners a third of the time, priced to lose a fifth of the money.

The efficiency rule won far less often. Read that again. The horse with the best late-to-early retention in the field won less than half as often as the favorite did. Its return was slightly worse than the favorite's. The paired comparison — race by race, the rule's return minus the favorite's return — ran against the rule, and its bootstrap confidence interval spanned zero, so the returns claim cannot even establish its direction. The duel settles what the interval leaves open. In the races where the rule's pick differed from the favorite, the favorite won far more often than the rule's horse did: a duel lost by a wide margin.

That is not a near miss. That is a ranking that disagrees with the market thousands of times and loses the disagreement by a margin of thousands.

The lab's frozen gates require all three of a positive paired interval, a positive duel, and a return above the baseline. R-PEI failed all three components. The verdict in the record, with its scope attached: **R-PEI NOT SUPPORTED on the research corpus.**

The control population agreed. On the larger control corpus, the same rule also posted a paired difference against it and lost its duel heavily, and was NOT SUPPORTED there too. When a verdict holds on the primary population and the control reads the same way, the lab records no cross-corpus dispute. The ratio did not fail because of which tracks were counted. It failed on both populations the lab keeps.

Why does the highest-retention horse win so rarely? The arithmetic of the ratio holds most of the answer, and you can see it without any statistics. A ratio rises two ways: the numerator grows, or the denominator shrinks. A closer who jogs early and finishes with an ordinary late figure posts a high ratio, because the early figure it is divided by is small. A brilliant front-runner whose late figure is strong in absolute terms but below its own enormous early figure posts a middling one. The ratio cannot tell those two horses apart, and it promotes the first kind on purpose. It selects, race after race, for horses whose defining trait is that they did little early. Some of them are finishers lying in wait. Most of them are just slow starters. The favorite, meanwhile, is usually a horse with absolute figures — early, late, or both — that tower over the field in plain terms. A ranking that divides one figure by

another throws away the size of both, and size is most of what the market was pricing. The duel count is what that trade looks like when the disagreements are settled by actual races.

The diagnostic that runs backwards

A defender of the ratio has a fair reply at this point, and the lab built the test for it before hearing the results. Fine, the reply goes: the ratio was never meant to pick long-priced closers out of the whole field. Use it on the favorite. A favorite with high retention is a favorite who will finish the job. A favorite with low retention is a speed horse living on borrowed time — the one to bet against. That is a diagnostic claim: an association between a favorite's own ratio and the favorite's chance of winning. It is weaker than a ranking claim, and weaker claims are easier to support. The lab measured it directly.

Split the favorites on the primary corpus into three groups by their own PEI_c, lowest third to highest third. A further group of favorites carried no ratio and sits outside the split. Now count wins.

The lowest-retention favorites won most often. The middle group won less often. The highest-retention favorites — the efficient ones, the keepers of their early speed — won least often. Their returns ran the same way: the lowest group lost the least, and the highest group lost more.

The association runs backwards. On the primary corpus, the favorites the ratio certifies as most efficient won least often, and the favorites it flags as wasteful won most often. The control tells the same story on the larger corpus: lowest third highest, middle lower, highest third lowest. Two populations, same

ordering. Whatever the ratio measures in a favorite, more of it did not mean a stronger favorite in these races. It meant a weaker one, by a clear margin of win rate from one end of the scale to the other.

Sit with what that does to the original story. The story said retention is strength. The count says that among favorites — the horses the market already respects — a high late-to-early ratio travels with the horses who win less. The plausible mechanism is the same denominator trick as before, now working inside the favorite population. A favorite whose early figure is modest relative to its late figure is often a favorite by default: the closer in a field with no real speed, the best of a weak lot, the horse the market backed because somebody had to be the favorite. A favorite with a towering early figure and a ratio dragged down by it is often a horse dominating its field at every call. The ratio reads the second horse as inefficient and the first horse as a model of economy. The races disagreed.

The caveat this chapter owes you

Now the discipline cuts the other way, against the chapter's own headline. The paragraph above reports an association in one population. It does not establish a law of racing, and this book will not print it as one. Read the caveat in full, because it is part of the result:

Efficiency terciles travel with price, field size, and race shape; the inverse pattern is an association in this population, not a law of racing. Favorites are not assigned to ratio groups at random. A high-retention favorite tends to arrive with a different price, in a different field size, in a different shape of

race — a field with little early speed produces both a closer-type favorite and a race where that favorite faces no pace pressure, or faces a lone front-runner who never gets caught for reasons that have nothing to do with ratios. Until those strands are separated — until high- and low-ratio favorites are compared at the same prices, in the same field sizes, in the same race shapes — the slide from the lowest third to the highest third describes who these favorites were, bundled with everything that traveled with them. It does not prove the ratio itself harmed any horse's chance, and it does not license a new rule in the opposite direction. Betting the *lowest*-retention favorite is not a system this chapter is selling you; that would be the same mistake with the sign flipped, a ranking claim dressed in a diagnostic's clothes.

What the count does settle is narrower, and it is enough. The burden of proof sat on the efficiency claim. Its own diagnostic, run in its own favored direction, on its own chosen subject — the favorite — came back pointing the other way, on the primary corpus and on the control. A claim does not have to be disproven as physics to be refused as handicapping. It has to fail to support itself. This one failed, and the failure is recorded with its caveat attached, so that nobody — including this lab, in a later season, with a convenient memory — upgrades a bundled association into a law.

Notice, too, what the diagnostic cannot rescue. Even if a future, cleaner study separated the strands and found a real retention effect hiding inside the slide, the ranking rule's record would stand untouched: a paired difference against it, an interval across zero, and a duel lost by a wide margin on the primary

corpus. Association, ranking, probability, and returns are four different claims. A repaired association would be a new starting point for a new frozen test. It would not rewrite this one.

The exact form: same ratio, our own cards

One objection remained, and it was a good one. The corpus test used a corpus form of the ratio — late figure over early figure, from the figures the research records carry. The form a player actually computes from a published past performance divides the late pace figure by the second-call figure: LP over E2. Different inputs, arguably a different measurement. If the corpus form failed because it was a rough stand-in, the exact form deserved its own count before the idea was closed.

That test ran as Phase 2 of the same program, on different evidence, and its evidence class is different, so this section carries its own label. The population is **OWN**: ThoroughWiz's own archived product cards, settled against the same days' results. Five race days — October 30 and 31 and November 2, 3, and 4, 2026 — produced 239 card races. Of those, 237 could be scored: one carried no final result, and in one the winner was not in the post-scratch field. The rule is the same rule with the exact figures: in each race, select the starter with the highest bris_lp divided by bris_e2. The baseline is the morning-line favorite in the same race. The lab labeled this study secondary evidence before it ran — small count, five days, the pile-in population of races the product happened to file rather than full cards — and no verdict in this chapter rests on it alone.

The favorite on those 237 races won 33.33% and returned –20.65%: the same favorite as everywhere else in this book, a third of the races, a fifth of the money. The exact-form efficiency rule won 12.24% of its races and returned –26.81%. The paired difference was –6.16 points. Its bootstrap 95% confidence interval ran from –49.86 to +53.01 points — an interval so wide it admits almost anything, which is exactly what 237 races buys you, and exactly why the study is labeled secondary. The duel, which does not depend on prices, was –50: in the 205 races where the rule's pick differed from the favorite, the rule's horse won 22 times and the favorite won 72 times.

The verdict: **R-PEI-BRIS NOT SUPPORTED**, as secondary evidence. The exact form reads the same way as the corpus form — a ranking by pace efficiency does not beat the morning line — and it reads that way at the same scale of failure: a win rate in the low teens against a baseline in the low thirties, and a duel lost by fifty in two hundred and five tries. Two forms, two populations, two evidence classes, one ordering. The corpus test, with its full count and its interval across zero, is the load-bearing wall. The product-card test is the confirmation that the wall is not an artifact of the corpus's substitute figures. Neither test leaves the efficiency ranking anywhere to stand.

Be careful, once more, about what the small test cannot do. Its interval is wide enough to hide a modest true effect in either direction, and a defender who says "237 races prove nothing" is right about the interval and wrong about the conclusion. The small test was never asked to carry the verdict. It was asked one question: does the exact form, on our own settled cards, show any sign of reversing the corpus result? A reversal would have shown up as a win rate near the favorite's and a duel near even.

It showed neither. Secondary evidence that agrees with primary evidence strengthens the record. That is all it does, and the record labels it so.

One more rescue was tried: efficiency, gated by pressure

A careful reader will think of the combination, because the lab thought of it too, and wrote it into the same program as the priority experiment before any results were read. Perhaps the ratio fails alone because it cannot tell a genuine finisher from a plodder. Gate it, then. First find the starters under real competitive pressure — rivals whose early figures crowd their own — and rank only those horses by efficiency. Or run it the other way: let efficiency speak only where the race shape gives a finisher a target to run at. The frozen form of the idea, R-PEI×CPE in the lab record, restricted the efficiency ranking to starters carrying enough pressure exposure, and fell back to the favorite where none qualified.

On the primary corpus, over the same evaluation races, the combination won far less often than the favorite did and returned less than the favorite did. Its paired difference against the favorite ran against it, with a confidence interval across zero. Its duel was lost by a wide margin, nearly as deep as the ungated rule's. On the control corpus it also posted a paired difference against it and lost its duel heavily. The verdict was NOT SUPPORTED on both populations the lab keeps, and the program's own rule — combine only after an individual method is supported — carried nothing forward from it.

Record what that closes off. The efficiency idea was not abandoned after one careless test. It was tested as a whole-field ranking, as a diagnostic on favorites, in a pressure-gated combination, and again in its exact published-figure form on our own settled cards. Four doors, all counted, all closed by their own numbers. Pressure exposure itself, as Chapter 14 will show, does describe a favorite's day honestly — favorites facing more matching early speed win less often. But describing the favorite's trouble is a different claim from naming the horse who profits from it, and gating a failed ranking behind an honest description did not turn the ranking into a working one. When an idea needs a rescue, the rescue gets its own frozen test and its own count. If the rescue fails too, the honest sentence is the short one: the idea was tried, in the forms its defenders asked for, and it did not support itself.

What to do with a ratio at the track

You will meet this ratio again. It circulates in software, in tip sheets, and in the quiet authority of the word *efficiency* itself. This chapter leaves you with the working rules the two tests bought.

When a number divides one figure by another, ask what it throws away. A ratio discards size on purpose. Sometimes size is noise and discarding it clarifies. In a horse race, size — how fast the horse actually ran, early and late, in figures the whole field shares — is most of the information the market prices. A measurement that cannot distinguish a slow closer from a fast front-runner will promote the slow closer every time their ratios tie, and it will do so with an engineer's confidence. Ask for the two figures separately before you accept their quotient.

When a measurement arrives with a story about machines, be most careful. The story is doing work the count has not done. Horses are not engines with a fuel tank of speed to spend wisely or wastefully. A horse that runs its late figure close to its early figure may be rationing, or may simply have been unhurried in a slow early pace, or may be a plodder whose early figure was never high enough to fall from. The ratio is silent on which. The story fills the silence with a picture of thrift. Pictures are not evidence.

And when you test an idea — yours, a book's, this book's — test the claim you actually intend to use. If you intend to bet the highest-ratio horse, count that rule against the favorite, in full, the way R-PEI was counted on the research corpus: a paired difference against it, an interval across zero, and a duel lost by a wide margin. If you intend only to shade your opinion of a favorite, count the diagnostic, the way the terciles were counted: the lowest-retention favorites won most often and the highest-retention favorites won least often, with the confounding caveat stapled to it — efficiency terciles travel with price, field size, and race shape, and the inverse pattern is an association in this population, not a law of racing. Do not let the second count borrow authority for the first claim, and do not let either borrow the authority of the word *efficiency*.

The myth, measured, comes down to this. On the largest count the lab could run, the most "efficient" horse in the field won far less often than the favorite did, and the most efficient favorites won less often than the least efficient ones. The ratio kept its engineering shine through both counts. Shine is not a claim. Counted claims are all this book has ever asked you to trust —

and the next chapter's stories, the lone-speed horse and its cousins, will need the same counting, because they sound even better than this one.

Chapter 12 — Lone speed and other stories that invert

Every handicapper knows this story. Tell it at any track, in any simulcast room, and heads nod.

A horse has the best early speed in the race. No other horse can run with it early. It will get the lead alone, set its own pace, and relax. Then, late, it will have nothing left, because its late figures are weak. The closers will come. The lone speed horse will stop. The favorite, if that horse is the favorite, is a false favorite. Bet against it.

The story sounds right. It has a shape. It has a mechanism. A horse that spends itself early pays late. A horse with no early rival controls the race until it does not. You can picture the whole race before it runs.

Our lab tested that story. The test did not go the way the story says.

This chapter is about that test, and about a second test that went half right and half wrong. Neither result is a trick. Both are ordinary. Stories that sound true fail in ordinary ways, and you need a plain method for telling the difference before you bet on one. That method is the real subject here. The two screens are just the worked examples.

Write the story down as a rule

A story you cannot write down as a rule cannot be tested. It can only be remembered when it works and forgotten when it does not.

So the lab wrote the lone-speed story down. The preregistration for the test was frozen before anyone read an outcome. That detail matters, and Chapter 4 explains why. A rule written after you see the results is not a test. It is a description of the results wearing a rule's clothes.

Here is the rule, in the form the lab filed it. The screen is called V1 in the lab record.

Flag a favorite when two things are both true:

1. The favorite is the unique early-speed maximum in its field, by at least 5 points on the lab's early pace figure.
2. Its late pace percentile in the field is below 40.

Read that slowly. The first condition is the lone speed part. The favorite does not just lead the field on early speed. It leads alone, and by a clear margin. No rival sits close enough to press it. The second condition is the weak late part. Among the horses in its own race, the favorite's late pace figure sits in the bottom part of the field. It starts fast, by the numbers, and finishes slow, by the numbers.

That is the story, translated. The horse gets loose. The horse has little left. The screen flags the favorite as vulnerable.

A screen is not a bet. Hold that distinction. A screen sorts races into two piles: flagged and unflagged. It makes one claim at first, an association claim. It says favorites in the flagged pile

behave differently from favorites in the unflagged pile. The lab then asks narrower questions. Do flagged favorites win less often? That is still association, stated as a win rate. Do they return less on win bets? That is a returns claim. Does a rule that bets against them, or bets the horse that beats them, beat the favorite baseline? That is a ranking and returns claim, and it needs its own test. This chapter grades the screens on the first claims only, because the screens never earned the later ones.

Association, ranking, probability, and returns are four different claims. This book grades them separately in every chapter. A screen can describe a group honestly and still give you no bet. A screen can also fail at the first step. V1 failed at the first step.

What the count showed

The population comes first, as it does in every chapter of this book. The lab states its population and its denominator in its report instead of filling gaps in. You should expect the same habit from any count you trust: the denominator is stated, and any missing race is named.

On the primary population, the favorite won at its usual rate and returned its usual loss on win bets. That is the baseline from Chapter 1. Keep it in view. Every group in this chapter gets compared with it, and with its own unflagged remainder.

V1 flagged very few favorites.

Written strictly, the lone-speed, weak-late pattern occurred only rarely. That rarity is itself a finding. A pattern this strict cannot carry a season, whatever its results say.

The flagged favorites won more often than the unflagged favorites. Not a little more. The direction was clear, and it was the wrong direction for the story. On returns, the flagged group lost at about the same rate as the unflagged group. It did not lose less. It did not collapse. It paid like an ordinary group of favorites while winning like a strong one, which means its winners paid short prices, as strong favorites do.

The lab's own sentence for this result is short, and this book adopts it: as specified, V1 does not identify false favorites.

That sentence is the verdict. A failed screen. Not a discovered strength. The rest of this section explains why you must not flip it over and use it the other way.

Why you cannot flip a failed screen

The temptation arrives at once. If flagged favorites win more often, why not bet them?

Because a very small group cannot support a strong claim in either direction. That is not caution for its own sake. It is arithmetic.

A group this small swings on ordinary luck. Change a few results — a few photos, a few bad breaks, a few rides — and the win rate moves a lot. The lab does not need a confidence interval to make the point. You can see the fragility in the rarity itself. A claim that rests on a handful of races rests on a handful of afternoons.

There is a second reason, and it matters more. The screen was built to test one direction. The preregistration named the pattern in advance: lone speed, weak late, vulnerable favorite.

The data answered that question. The answer was no. Inventing the opposite claim after the answer arrives — "lone speed favorites are actually strong, bet them" — is a new claim, not a result. It has not been tested. It was selected *because* it fit those few races. Of course it fits them. It was built from them.

This is the multiple-testing trap from Chapter 4, in its smallest and most tempting form. Run enough screens, and some small group will look special by chance. If you keep the groups that look special and discard the rest, you have not found anything. You have sorted luck and kept the top of the pile.

The control count made the same point from the other side. On the control population, V1 again flagged a small group. Again the flagged group won more often than average. Again it returned less, because its winners paid less. Two populations, same shape: the win rate points one way, the returns point the other, and the group stays small in both. Nothing in either count turns the screen into a bet, in either direction.

So the honest record holds three facts at once. The pattern is rare. The favorites it flags win often. They still lose money at about the normal favorite rate. All three facts come from the same tables. A writer who reports only the middle fact sells you a story. A writer who reports only the first fact buries one. This book reports the count in full or not at all.

Why the story felt so true

A failed test does not mean the story was foolish. It means the story, written as a strict rule, did not describe the population. It helps to ask why a plausible story missed, because the answer teaches you how to read the next one.

Start with the rarity. The strict rule needs a favorite who leads the field on early speed alone, by a clear margin, *and* sits low in the field on late pace. Horses like that exist. Favorites like that are rare, for a reason the story itself supplies. The crowd prices a horse partly on how its race sets up. A favorite with a clear early edge has a visible advantage. Bettors see it. The morning line sees it. The price reflects it. The pattern the screen hunts is, in part, already in the number the horse carries to the gate.

Next, consider what "weak late" means in the rule. The lab measured late pace as a percentile inside the horse's own field. Low in that field does not mean the horse stops. It does not mean the horse cannot finish. A strong horse in a field of closers can post a modest late percentile and still hold on, because holding on is a comparison, not an absolute. It only needs to finish better than the horses chasing it, on that day, over that distance.

Then consider the lead itself. The story treats a lone lead as a debt the horse pays late. A lone lead can also be a gift. A horse alone in front chooses its own pace. It is not pressed. It does not fight for position. It saves the energy that a duel would spend. Which effect wins — the weak late figure or the soft trip — is an empirical question. Stories cannot settle it. Counts settle it, and the count said the flagged favorites won more often than the average favorite.

None of this proves lone speed is good. Reread the last section if that sentence surprises you. A rare pattern proves nothing strong in either direction. What the test shows is narrower and more useful: this story, in this strict form, on this population, did not find the weak favorites it promised. A handicapper who

bets against every lone-speed favorite on the strength of the story alone bets against a group that, in the only full count the lab has, won often.

The second story: the closer who leaves too much

The lab ran a second vulnerability screen in the same preregistered test. It tells the opposite half of the lesson. V1 failed in one direction, cleanly. V2 half worked, and the half that failed is the half that pays.

The story goes like this. Some favorites do their best running late. They close. Closing is a fine way to win a race, until the race leaves the closer too much to do. If a favorite's early pace is slow relative to its field, it starts far back. Its late run must pass every horse, make up every length, and arrive in time. Against a big field, that task gets harder. More traffic. More ground to make up. More chances for the trip to go wrong. A favorite like that, the story says, is vulnerable.

The lab wrote that story as a rule too. The screen is V2.

Flag a favorite when three things are all true:

1. The field has at least 9 starters.
2. The favorite's late pace percentile in the field is at least 90.
3. The favorite's early pace figure sits in the bottom half of the field's early pace carriers.

That is a strong closer, in a large field, who starts slowly by the field's own standard. The closer who leaves too much.

On the primary population, V2 flagged a workable number of favorites. Not a large group, but large enough that a few photos do not rewrite it.

The flagged favorites won less often than the unflagged favorites, and they returned less on win bets.

On the primary population, the screen separates in the direction the story predicts, on both measures. Flagged favorites win less often. Flagged favorites lose money faster. If the lab had only one population, this chapter would report a screen that describes a genuinely weaker group of favorites, and then — carefully, under the discipline of this book — it would still stop short of a bet, because a screen that finds weak favorites does not by itself name the horse who beats them. Remember Chapter 14's rule, stated early: describing a favorite's weakness never, by itself, names its conqueror. But the description, on the primary population, holds.

The lab has two populations. That is why it has two.

On the control population, the win-rate separation held. Flagged closers won less often there too.

Their returns reversed. The group that won less often lost *less* money on the control, because the closers who did win paid prices that covered more of the losses. The returns claim — the claim a bettor can spend — points the opposite way on the control population from the way it points on the primary population.

The lab records V2 exactly that way, and this book keeps its wording. V2 is a diagnostic screen, not a ranking rule. It carries no supported verdict. It separates win rate on both populations and returns on the primary population only. In one word the lab's summary uses and this chapter adopts: inconsistent.

What inconsistent costs you

Inconsistent is not a shrug. It is a specific and expensive condition, and a casual player should learn to price it.

Suppose you read only the primary result. You build a habit: bet against the strong closer in the big field, or pass its races, or spread your money on the horses in front of it. On the primary population that habit matches the record. On the control population — a larger population, the same year, more tracks — the returns run the other way. A bettor who acted on the first count alone would act on a result that a second count, already sitting in the same lab file, contradicts on the measure that matters most.

This is why the lab runs a control at all. The control is not a second chance to shop for a better answer. It is a check on whether the first answer survives a change of population. A finding that depends on which list of tracks you count has told you something about the list, not yet about horses.

Notice, too, which measure reversed. The win rate held on both populations. The returns did not. That split is the four-claims discipline doing its work. Association held: flagged closers win less often, in both populations. Returns did not hold: the money result changed direction between populations. If the chapter had graded V2 on win rate alone and called it a wagering

finding, it would have printed a claim its own source file refutes. A horse can win less often and still be the better bet, if it pays more when it wins. Prices are part of every returns claim. There is no way around them, and screens that ignore prices will keep producing this exact disappointment.

So V2 ends the chapter in a different place from V1, and the difference is the lesson. V1 is a failed screen: it pointed the wrong way on its first claim, on a count too small to rescue. V2 is an inconsistent screen: its descriptive claim survived two populations and its money claim did not. Neither screen gives you a bet. They fail differently, and you read them differently. The first you discard. The second you file as a description — big-field favorites who close from far back win less often — and you refuse to spend it as a prediction about money until a test that names actual bets survives both populations.

How to test a story yourself

You will hear a new story next week. Racing produces them the way weather produces rain. A friend will have one. A podcast will have one. Your own eye will offer you one after a single afternoon of watching favorites stop in the stretch. Chapter 19 gives the lab standard as a kitchen-table method in full. This chapter gives the short version, built from the two failures above, because these two screens are the template.

Write the rule before you count. Every condition, every cutoff, every exclusion. V1 and V2 each fit in three short lines. If your story cannot fit in three short lines, it is not a rule yet. It is a mood. Moods remember their wins.

Name the population and the denominator. Corpus, year, tracks, races. V1's whole meaning lives in its denominator: it fired rarely in a full population of racing. A story that fires rarely cannot carry a season, whatever its win rate says. Count how often the pattern occurs before you count how it performs.

Compare against the unflagged remainder, not against your hopes. The flagged group means nothing alone. A high win rate sounds strong until you set it beside the unflagged group and see a group that wins more and loses the same. Every screen in this book prints both piles in its lab record. Demand both piles from anyone who shows you a pattern.

Separate the claims. Ask, in order: does the flagged group behave differently? Does it win less or more often? Does it return less or more? Does a stated bet on stated horses beat the favorite on the same races? V1 died at the second question. V2 passed the second, split the third, and never reached the fourth. Most stories die early. Let them.

Run the second population before you believe the first. If you keep records at two tracks, test at both. If a pattern works at one and reverses at the other, you hold an inconsistent screen, and the correct price for an inconsistent screen is zero. Do not average the two results into a small positive and bet that. The reversal is the information.

Refuse the flip. When a screen fails in the direction you aimed it, the opposite bet is a new, untested claim, however obvious it looks in the table that just failed. Write it down as its own rule. Test it on races the first test did not use to invent it. Until then it is a story about a few races.

Let small counts stay small. A rare pattern could not support a strong claim in either direction. Neither can the few races in your notebook from last month. This is not an insult to your notebook. It is the arithmetic of a game where the best single signal in the sport — the favorite — needs many races to pin down. Your pattern does not get a smaller standard because it is yours.

The point of the two stories

Neither screen in this chapter was foolish. Lone speed with a weak finish is a real shape; some of those horses do stop, in full view, every week. A closer who starts last in a big field faces a real task; some of those favorites do flatten out with too much ground left. Your eye is not lying when it shows you those races. Your eye is sampling. It shows you the races that fit the story, at the moment they fit, and it does not show you the denominator.

The denominator is the whole discipline. A rare flag in a full population. A workable flag count, two populations, one reversal. Once you see a story with its denominator attached, you cannot unsee it, and most stories get quieter. A few get louder. The next chapter takes up a pattern that is not rare at all — it covers a large share of every field you will ever handicap — and shows what a full denominator does to the most natural story of all: the horse with the best single number in the race.

Chapter 13 — One-factor wonders

Open any race on any card and sort the horses by one number. Sort by the speed figure. Sort by early pace. Sort by the handicap weight our lab prints, or by the draw statistic, or by the morning line itself. At the top of your sorted list sits a horse with the best single number in the race.

That horse pulls at you. It should. You did work to find it. The number is real. The horse really did run that fast, or start that quickly, or earn that weight, or draw that well, by the measure you chose. The story writes itself in one sentence: this horse owns the best figure in the race, and the crowd has it priced as an outsider.

Our lab has a name for that horse when it sits outside the market's leading group: a one-hit wonder. Not as an insult. As a counting category, defined in advance, so the lab could count every one of them in a full year of racing and ask the only questions that matter. How many are there? How often do they win? What do they pay?

The answers are the chapter. A large share of all starters are one-hit wonders by the lab's definition. As a group they win rarely. They return less on win bets than the favorites and the leading group do. And they supply a meaningful share of all winners — which is the fact that keeps the story alive, and the fact this chapter will teach you to hold at the right size.

The counting rule, stated exactly

A category you cannot define exactly cannot be counted honestly. Change the definition after you see the results and you can make the group look like anything. So here is the lab's definition, stated in full, before any result that depends on it. The lab filed this measurement on 2026-11-05 and changed nothing in its product because of it. It is measurement only.

First, the five factors. The lab's selection method, the consensus described in Chapters 1 and 10, scores a field on five factors: the morning line (ML), the handicap weight (WGT), the speed figure (SPD), early pace (EARLY), and the draw (DRAW). Each factor ranks the full field. Under the production construction, the top three horses in each factor receive competition points — 5 points for first, 2 for second, 1 for third. A horse with no value in a factor receives 0 in that factor. Ties share a competition rank, the way tied runners share a placing.

Second, the top four. The morning line's top four means the first four starters in ML-BASE order: morning line ascending, scratches removed, ties broken by program number, a starter with no morning line placed last. This is the group the lab's selection method lets compete for its pick. Horses outside it can score all the factor points they like. Those points strand. The selection comes from inside the four.

Now the definition.

A one-hit wonder is a starter outside the ML top four that holds a full-field top-three score — the 5, 2, or 1 competition points — in at least one of the five §26 factors (ML, weight, speed, early pace, draw).

Each starter counts once in the headline group, no matter how many factors it tops. A horse that leads the field in both speed and early pace is one wonder, not two. The per-factor counts later in this chapter work the other way, and the chapter flags it where it happens: in the by-factor discussion, a horse with two hits counts in both of its factor rows, so those rows overlap and add up to more than the group. The headline count never double-counts a horse.

One clause of the definition needs a plain explanation, because it looks like a contradiction. How can a horse outside the ML top four hold top-three points in the ML factor itself? Through ties. Competition ranking gives every horse at a tied price the same rank. The top four, though, has only four slots, and it breaks ties by program number. So a horse tied at a price whose competition rank is third can sit fifth or sixth in ML-BASE order — outside the four, holding an ML point. It is a wonder by the definition, on the ML factor alone. The group formed by that tie case is small, and the lab reports it separately so the tie case never hides inside the larger factors.

That is the whole rule. No judgment calls. No "the figure looked strong." A starter is in the group or out of it by arithmetic, and a reader with the same records gets the same group.

The population

The lab counted a full population of racing, with the funnel matching the rest of this book: races where the lab's selection started. Returns are $2 win returns.

Inside that population, the ML top four covers a little over half the starters. The outsiders — every starter outside the top four — make up the rest. The wonders and the no-hit outsiders divide that outside group between them.

Keep those groups in view. Every comparison in this chapter names which group it uses.

A large share of every field

One-hit wonders make up a large share of all starters.

They are not a curiosity you hunt for. They are the texture of every race. Nearly every race contains at least one wonder. Open tomorrow's card at random. Pick a race. The odds are high that at least one outsider in it tops some factor list you can build. In most races, several do. The pattern you "found" in this morning's past performances was there before you looked, in nearly every race, whether you looked or not.

This is the fact that reorganizes the chapter, and it deserves a pause. A signal that fires on a large share of all starters is not a discovery about a horse. It is a property of lists. Five factors, three scoring places in each, fields of eight or so horses: the arithmetic guarantees a wide spread of top-three finishes among outsiders. Somebody outside the top four will post a leading speed figure, because the top four cannot hold all the scoring places in a five-factor table. The wonder is not hiding. The wonder is what a field looks like when you sort it five ways.

So the question was never whether one-hit wonders exist, or whether you can find one. You can find one in almost every race, in under a minute, with any single column of figures. The

question is what the hit is worth. Worth has three measures in this book, and the chapter takes them in the book's order: association first, then the win rate, then the money.

What a wonder wins, and what it pays

Here is the headline finding from the lab's file, stated qualitatively.

The ML top four wins far more often than any outside group, and it supplies most winners.

One-hit wonders win rarely — about one race in fifteen, in the lab's phrasing — and they return less on win bets than the top four does. Wonders supply a meaningful share of all winners: roughly one winner in six.

Outsiders with no factor hit at all win still less often, return much less, and supply only a small share of winners.

All outsiders together win less often than the top four and return less.

Read the finding in layers, because it says two true things that pull in opposite directions, and a casual player needs both.

The first layer compares the wonder with its own class. Against outsiders with no hit — the honest baseline, horses the market placed outside its leading four and no factor rescued — the single hit raises the win rate substantially. It cuts the loss rate substantially. The hit is not nothing. A top-three score in a real factor carries real information about a horse's chance. If this chapter told you to ignore every factor list and bet outsiders blind, it would be giving you worse advice than the wonder

gives you. The association is real, it is large by the standards of this book, and it survives the control population in the next section.

The second layer compares the wonder with the group it wants to join. Wonders win rarely. The top four wins far more often, horse for horse. And on money, the wonder loses *faster* than the top four. The hit lifts the outsider well above the no-hit floor and still leaves it below the leading group on the measure a bettor spends. An improvement in association — the hit tracks winning — did not become an improvement in returns. Ranking and returns are different claims. This finding is that sentence with a full population behind it.

Now the third fact, the one that keeps the story alive at every track. Wonders supply a meaningful share of all winners. That is not a rare event you might never see. At a typical card, the arithmetic expects a wonder winner once or twice a day. You will see them. You will see them on days you bet them and days you did not. Some will pay well. Each one will feel like proof that the single number was the key to the race.

Hold that feeling up against the full count. Many wonder starters lost for each wonder winner. The winners are the ones the story remembers, because winners get circled, posted, and retold. The losers were in the same races, holding the same kind of hit, and they left no story behind. A method that shows you the winners and hides the full group of starters is not lying to you. It is sampling for you, the way your eye samples at the track, and the denominator is the part it leaves out.

Which factor, and how many hits

The by-factor breakdown carries the overlap warning from the definition: these rows overlap. A horse with a speed hit and a draw hit counts in both rows. The rows describe factors, not separate horses, and they do not add up to the headline group.

A draw hit is the most common, and its return is the least negative among the large factors. Read it with the overlap in mind and the caution this book applies to any subgroup: the draw row shares many horses with the other rows, it was not a preregistered bet, and no rule that bets the draw-hit outsiders was tested as a ranking rule in this measurement. The file reports a description. This chapter reports it as one.

A speed hit gives the highest win rate among the large factors. A weight hit sits in the middle. An early pace hit gives the weakest return of the four large factors, a fact worth setting beside Chapter 12, where early-speed stories also read better in the telling than in the count. The ML tie case stands apart, as the definition section promised, and it is small. The lab's own verdict on that cell is in its file and this chapter keeps it: the cell is real and too small to price. A small group cannot carry a returns claim, and no reader should carry one away.

Hit count tells a calmer version of the same story. Horses with exactly one hit win less often and return less than horses with two or more hits. More hits help, the way you would expect if each factor carries a little information. They help by a modest amount, and they land the multi-hit outsider at roughly the same return as the single-hit one — still below the top four's return, still winning rarely. A horse that tops two lists is a better outsider. The count does not make it a good bet, and it does not

close the gap to the group the market put in front of it. Complexity bought a little association and no returns — Chapter 10's lesson, measured again from the other side.

The checks: a second definition and a second population

A measurement this central gets checked twice in the lab's file, and both checks belong in the chapter, because a number you cannot check is a number you are asked to take on trust.

The first check changes the top four. The headline uses the ML-BASE four over starters, scratches removed before ordering. An alternative construction builds its four from the filed field, scratches not removed first. Under that definition the wonders grow, as they must when scratched horses vacate slots. Every comparison lands in the same place as the headline. Same win-rate lift over the no-hit baseline, same return short of the top four, same shape of winners' share, larger.

The second check changes the population. On the control population, the wonder share, the win rate, the return, and the winners' shares all match the primary finding in shape. No-hit outsiders again win rarely and return much less. The top four again wins most often and supplies most winners. The lab's conclusion is one line: the headline is not a track-universe artifact. Under this book's standard amendment on scope, that agreement matters: a gate-relevant verdict that flipped between populations would go back to the reader as an open question. This one does not flip. On both populations, the wonder is a large share of the starters, wins rarely, lifts well clear of the no-hit floor, and returns less than the leading group.

Why the best single number feels like more than it is

Step back from the tables and ask why this particular story survives every count that refutes its strong form. Three reasons, and each one is a habit you can catch in yourself.

The first is the size of the group. A pattern that covers a large share of starters will produce winners constantly, forever, whether it has any value or not. Frequency masquerades as confirmation. You do not need a good memory to collect wonder winners. You need only to keep handicapping. They arrive on schedule.

The second is the real association. This story is not a mirage like a screen that points the wrong way. The hit genuinely raises an outsider's win rate against the no-hit baseline. A handicapper who senses that the horse with the field's best speed figure is *live* senses something true. The error is not in the sensing. The error is in the step from "more likely than the other outsiders" to "likely," and from "likely" to "worth the price." A low win rate is more likely than a still lower one and still means the horse loses most of the time. The crowd, which priced the horse outside its top four, had already charged for the hit. That is why the return stays negative instead of somewhere flattering: the price knew.

The third is the missing comparison. At the track, the wonder gets compared with the bet you almost made, or with the favorite who lost, never with the full group of wonder starters. Against one beaten favorite, the wonder looks like insight. Against its own class — every outsider with a hit, all year — it looks like what it is: the better half of a losing population. Every test in this book is, at bottom, a device for forcing the full

comparison. Build the same device for your own patterns. When a single-factor horse wins for you, write down the factor and the position. Then, for the next month, write down *every* starter that fits the same description, winner or loser, bet or no bet. Count them all at the end. You have just run the kitchen-table version of this chapter, and your count, not your memory, gets the verdict.

What to do with a one-factor horse

This book does not end a measurement chapter with a ban. It ends with the claim stated at its correct size, so you can use what is true in it and refuse what is not.

A top-three factor score in an outsider is information. It separates that horse from the no-hit outsiders, on association and on returns, in two populations and under two definitions of the top four. Use it the way the lab's own method uses it: as one vote among several, inside a group the market already respects. The lab's consensus counts factor points only inside the ML top four — not because outsiders never win, but because this measurement, and the ones around it, show what a stranded point is worth on its own. One vote does not elect.

Refuse the strong form of the story. The horse with the best single number in the race is usually just a horse — one of several such horses in every field, in nearly every race, winning rarely as a group. When the number tempts you, ask the chapter's three questions in order. How many starters fit this description, here and in general? What do they win as a group,

not as a memory? What do they return? If you cannot answer the first question, you do not have a pattern yet. You have a horse you like, with a number attached.

One worked picture fixes the scale before you go. Take a single race with eight starters. The top four take four places. Four outsiders remain. On the lab's rates, expect some of those outsiders to hold a factor hit. Expect the top four, as a group, to win that race most of the time over the long run. Expect a wonder to win it sometimes, regularly enough that you remember it. And expect the no-hit outsider to win it rarely. Now run that race many times in your head, with different horses and the same structure. The wonder wins often enough that you remember its name. It loses often enough that betting it every time costs you money. Both expectations come from the same finding. A handicapper who holds only the first one owns half a fact, and half a fact at the track costs the same as a whole error.

The next chapter turns to the two measurements in the lab's program that described their subject honestly and still produced no bet — pressure on a favorite, and clear separation from the field. They are the strongest descriptions in this part of the book. They earn a chapter precisely because they show where description ends, even when the description is true.

Chapter 14 — Pressure and separation: two measurements that tell the truth

Most handicapping talk about a favorite is talk about the favorite. Is it fast enough? Is it in form? Is the class right, is the trainer hot, is the distance suitable? The horse gets examined the way you would examine a used car, panel by panel, as if its afternoon depended on its own qualities alone.

It does not. A race is not an examination. It is an argument, held at speed, and a favorite's afternoon depends at least as much on who argues back as on anything printed in its own past performances. Two measurements in our laboratory's records make that plain, and they make it plain in a way most handicapping measurements do not: by surviving a full count of races with their story intact. Both measurements are about the favorite. Neither one is about the favorite's own figures. One counts the company pressing it early. The other measures how far the field's best speed figure stands from the rest.

This chapter is about those two measurements, and it comes with a discipline attached, stated now so nothing later can be misread. Both measurements *describe* favorites honestly. The laboratory then did the obvious next thing — it built pick rules on top of each description and tried to turn description into prediction — and both rules failed their tests, in the same report, on the same races. The chapter carries both halves. A measurement that tells the truth about a race is a valuable thing for a casual player to own. It becomes a harmful thing the moment it is sold, by a book or by your own hopes, as a way to

name a different winner. Those are different claims, and this book grades them separately: association is not ranking, ranking is not probability, and none of the three is a return.

The first measurement: who else wants the front

Start with the simplest question you can ask about a favorite that likes to run early: how many other horses in this field can run early with it?

That is the whole of the first measurement. The laboratory calls it competitive pressure exposure. Take the favorite's early pace figure — a number that summarizes how fast the horse customarily runs in the opening part of its races. Then count its rivals whose early pace figure is at least as high, allowing a small margin: a rival counts as pressure if its early figure comes within a few points of the favorite's, or beats it. The count is the measurement. Zero means the favorite can expect the early running largely to itself. Three or more means the first part of the race will be contested by a crowd.

The margin matters, and the laboratory fixed it before looking at any outcomes, the way Chapter 4 requires. The main setting allowed a rival within five points to count as pressure. Two other settings — within three points, and within eight — were written down at the same time as sensitivities: reported alongside, never shopped for afterward. Keep that in mind, because the three settings do not tell identical stories, and the differences are part of the lesson.

Here is the qualitative finding. At the tightest margin, the picture is as clean as anything in this book. Favorites with no comparable early-speed rival won much more often than

average. Favorites with one or two such rivals won more often than average. Favorites with three or more — the great bulk of all favorites — won less often than average.

Read that slowly. A favorite that will be left alone early wins notably more often than a typical favorite. A favorite that will be pressed by several rivals wins less often. Nothing about the favorite's own quality changed across those groups. Its figure is the same figure. What changed is the company. Pressure is not a metaphor. Counted over a full population, it moves the win rate materially from the quietest group to the most crowded one.

The other settings qualify the picture without rescuing anyone from it. At the preregistered main margin, the bands come out less tidy, and the top bands swap places on small counts, which is what small counts do. At the widest margin the no-pressure band nearly vanishes, because a margin that wide counts almost every field as pressure. The honest summary is not "the measurement works at one setting." The honest summary is that the direction is the same at every setting the laboratory wrote down in advance: the more comparable early speed a favorite faces, the less often it wins, and the size of the effect depends on how you draw the counting line. A measurement whose answer moves when you move its ruler is telling you something true and imprecise at the same time. Both halves belong in your notebook.

Two more facts belong with the finding, because a casual player deserves the whole picture and not the flattering rows. First, some favorites in the population carried no early pace figure at all and sit outside these bands entirely; the measurement says nothing about them, and a method that is silent about some favorites should say so. Second, even the quietest group lost

money. The no-pressure favorites returned a loss on win bets: better than the favorite's overall return, better than the pressed groups, and still a loss. Hold that fact. It is the hinge of this chapter, and we will come back to it.

The second measurement: how far ahead the best figure stands

The second measurement asks not about the favorite but about the shape of the field's speed figures. Every starter carries a figure the laboratory built from its recent races — the median of its last three, a typical-day number rather than a best-day one. In each race, take the highest of those figures and subtract the second-highest. The difference is the field's separation: how far the best-figured horse stands above its nearest rival on paper.

Split the races into thirds by that gap and ask how the favorite did in each third. In the bottom third — fields where the top two figures sit close together — favorites won less often than average. In the middle third, they won at about an average rate. In the top third, where the best figure stands well clear of the second-best, favorites won much more often.

The same pattern, in the same direction, at nearly the same size, appears on the control population. That is a control, not an independent replication; it is the same underlying records with the track filter removed, and this book's evidence key reserves "independently replicated" for a genuinely separate dataset. What the control does rule out is the suspicion that the pattern is an artifact of which tracks the product happens to cover. It is not. It is a property of the records.

A second cut at the same measurement agrees. Line up the races by whether the favorite won or lost, and look at the separation in each pile. In races the favorite won, the typical gap between the top two figures was larger. In races the favorite lost, the gap was smaller. The gaps overlap heavily; plenty of favorites win bunched fields and plenty lose runaway ones. But the center of the distribution moves the way the thirds say it moves. When one horse's typical figure stands well clear of the field, the favorite's afternoon improves with it.

Notice what both measurements have in common, because it is why this chapter exists. Neither one looks inside the favorite and pronounces on its soul. Pressure counts the rivals. Separation measures the field's shape. Both are statements about a *situation*, and situations, it turns out, are what a favorite's record is mostly made of. The crowd prices the horse. A striking amount of the outcome is priced into the situation instead — or rather, not priced, because the crowd's price on the favorite barely moves between these groups, as the returns in a moment will show.

What these measurements claim, exactly

It is time to be precise, because everything useful in this chapter depends on keeping four claims apart.

The pressure and separation results are **association** results. They say: in these races, favorites in one kind of situation won at one rate, and favorites in another kind won at another rate, and the difference is far too large and too orderly to be a counting accident. That claim is solid. It is also all they say.

They are not **ranking** results. A ranking claim would say: use this measurement to reorder the field and you will name winners more often than the morning line's order names them. Nothing in a table of favorite win rates by situation says anything about which other horse wins when the pressed favorite loses. Pressed favorites lose most of their races. The measurement does not tell you whether the winner is the main pressure horse, a stalker sitting third, or a closer who never saw the early fight. It cannot. It never looked.

They are not **probability** results either, not in the strong sense. It is tempting to read a group win rate as "this favorite, today, has that chance." That reading smuggles in assumptions the count does not support: that today's race belongs to the same population, that the group rate applies to this member of the group, that nothing else about this race — its distance, its surface, its field size, the favorite's price — moves the number. Group rates are honest descriptions of groups. A probability claim about one race is a further step, and it has to be tested as one: state the chances in advance, then check over many races whether horses you called at one rate won at that rate. The laboratory did not run that test on these measurements, and this chapter does not pretend it did.

And they are not **returns** results. The returns sit beside the win rates, and they refuse to cooperate with any exciting reading. Every group in both measurements lost money on win bets, including the best group of each. If these measurements handed you an edge, the crowd — which sets the prices the returns are computed from — would have had to leave that edge lying in plain sight, in a table anyone can build from figures anyone can buy. It did not. The crowd prices pressed favorites and clear-

figure favorites only slightly differently from other favorites, and "slightly" turns out to be almost exactly enough. Description, again: the situations differ materially in win rate and by a smaller amount in return, and a smaller loss against a takeout tide is a smaller loss, not an income.

The rules the laboratory built on them — and what happened

A laboratory that stopped at description would be a library. The point of the place is to test whether a description can be promoted, and promotion has a price of admission, written down before the races were counted. In this program the price was three gates, all judged on the primary population: a new pick rule had to beat the morning-line favorite on the same races by a paired margin whose statistical interval sat entirely above zero; it had to win the head-to-head duel — in races where its pick differed from the favorite, its pick had to win more often than the favorite did; and its return on win bets had to finish above the favorite's return. Pass all three and the rule is supported. Fail any one and the record says not supported, and nothing is combined, adopted or sold.

Two rules were built, one on each measurement.

The pressure rule reasoned the way a storyteller would. If pressure cooks the favorite, bet the horse the pressure helps: among starters who are not themselves the field's fastest early horse and whose own pressure count is at least two, take the one with the strongest late figure — the stalker or closer best placed to pick up the pieces. If no horse qualified, the rule kept the favorite. The rule has a name in the lab's files and a certain

elegance, and none of that survived contact with the count. Its picks won much less often than the favorite. Its return sat below the favorite's return. Its paired margin against the favorite was negative, with an interval that crossed zero. And the duel was a rout: in races where the rule named a different horse than the favorite, the favorite won far more often than the rule's horse did. Not supported, on every gate, and the sensitivity settings of the pressure margin read the same way.

The separation rule reasoned the way an engineer would. Four measurements were in play — the median speed figure, the early figure, the late figure, and a late-to-early ratio — so let each race's pick be the horse that leads the most of the four, with ties broken by the size of its leads and then by the morning line's order. This rule came closer than anything else in its program, and its near-miss is the most instructive result in this chapter. Its return finished above the favorite's return, and its paired margin was positive — the only positive one in the program. Then the other two gates spoke. The interval around that margin included zero, and a margin that might be zero has not been shown to exist. And the duel was clearly negative. In races where this rule disagreed with the favorite, its horse won far less often than the favorite did. Read those two facts together and the shape of the near-miss explains itself: the rule's edge in return came from the *prices* of the winners it happened to catch, not from catching more winners. It named the winner far less often than the favorite did in disputed races, and got paid better when it was right. A return built that way, on an interval that spans zero, is a description of some races that already happened. It is not a prediction about the next ones. Not supported.

Sit with the pattern, because it is the finding of this chapter and, in a way, of the whole research program these measurements came from. Pressure genuinely describes which favorites are vulnerable. Separation genuinely describes which favorites are strong. And when the laboratory asked either measurement the only question a bettor actually needs answered — *then who wins instead?* — both went silent, and the rules built to speak for them lost to the favorite they had so accurately diagnosed. Naming the favorite's weakness is not the same act as naming its conqueror. The horses that beat a cooked favorite are a different population from the horses a plausible story points to, and a full count says the story points wrong.

What would promote a description to a prediction

Since this book asks you to grade claims, it owes you the promotion standard in plain terms. A description like the ones above earns the right to be called a prediction only by surviving, in order:

A rule, written in full, in advance. Not "pressure matters" but an exact procedure that, handed a field before the race, names one horse or names a pass — including its margins and tie-breaks, frozen before outcomes are read. Every adjustment after seeing results restarts the clock, because a rule tuned on the races that judge it has not been tested at all. Note what this does to the pressure bands: one setting produced the cleanest table, and choosing it *because* it produced the cleanest table would be exactly the sin this rule exists to prevent. The laboratory reported it as a sensitivity and built its rule on the setting fixed in advance. The clean table is a description. It is not a license.

A baseline beaten on the same races. The comparison is never the rule against your memory of favorites, or against favorites in general. It is the rule against the plain morning-line favorite, race by race, on identical races, with the denominator stated — the paired comparison this book has used since Chapter 1.

An interval that clears zero, and a duel that is won. A point estimate is a rumor about an effect; the interval is the effect's honest size range, and if it spans zero the effect may not exist. The duel guards a different flank: a rule can post a pretty return on a handful of long-priced winners while losing the winner-naming contest badly, as the separation rule did. Both guards exist because each has caught a failure the other would have waved through.

Races the rule's builders had not seen. A backtest, however honest, is exploratory: the builders knew the population, chose the measurements, and wrote the gates with the data in the building. The final promotion step is a frozen rule meeting future races and keeping its record there. The measurements in this chapter have not taken that step, because the rules built on them failed earlier steps. The descriptions themselves remain what they were on the first count: true of the population counted, on the stated tracks, in the stated period.

Anything less than that full climb is not a failed prediction. It is a description wearing a prediction's clothes, and the difference is worth money, because bets placed on descriptions are priced by a crowd that has already read them.

How a casual player uses a description

If these measurements cannot name a pick and cannot make a bet, what are they for? More than nothing, and less than a system. They are for reading a race with your expectations set correctly, and correct expectations are a real, if unglamorous, form of money management.

Use pressure before you use anything else in a race, as a *temperature check on the favorite*. You do not need the laboratory's figures to do a fair version of this. Open the past performances and ask the counting question: how many horses in this field habitually run in the first flight early, at speeds comparable to the favorite's? If the answer is none — the favorite projects to an uncontested lead — then treat the favorite as the strong favorite its situation says it is: in the counted races, that horse won more often than a typical favorite. If the answer is three or more, treat the favorite as an ordinary-to-weak one, whatever its form looks like in isolation, because in the counted races that horse won less often — and know that the race is now a race where *someone* benefits from the fight, without the count telling you who.

Use separation the same way, as a check on how much daylight the figures show. When one horse's typical speed figure stands well clear of the field, the race is more likely to run to paper, and the favorite's afternoon improves. When the top figures are bunched within a couple of points, the race is the kind where paper stops helping early: form angles, trips and luck decide more of it, and the favorite's edge over the field is thinner than its price will usually admit.

What do you *do* with a corrected expectation? Three honest things. First, you can pass with better aim: a pressed favorite in a bunched field, at a short price, is the profile of the most reliably disappointing bet in the sport, and declining it is a decision with a measured basis rather than a feeling. Second, you can demand more price: if you like a horse against a favorite whose situation is strong — uncontested, clear on figures — the count says you are fighting a stronger favorite, and your horse's odds should pay you for that fight before you take it. Third, you can stop being surprised, which sounds small and is not. Most casual players' worst sessions start with a "how did that lose?" favorite and continue with chasing. A favorite cooked by pressure did not betray you. It ran the race its situation described. Players who expect the situation lose less often to it, because they sized and chose their bets as if it were real.

And three things you must not do, each of which this chapter's own findings will tempt you toward. Do not bet the closer because the favorite is pressed — that is the pressure rule, and it was not supported. Do not bet the horse that "leads the most lists" — that is the separation rule, and it lost its duel clearly. And do not bet the quiet favorite because its group won often — that group's return was still negative, because the crowd had already priced the quiet. The measurements tell you the truth about the race. The truth about the race was in the price before you got there. Both things are real, and holding both at once is most of what this book is trying to teach.

The three lessons of pressure and separation

First: a favorite is a situation, not just a horse. Count its early company and measure its field's separation before you judge its form. On a full count of races, those two situational facts moved the favorite's win rate materially, more than most of the horse-level angles in this book ever moved anything.

Second: description is not prediction, and the gap between them is where systems are sold. Both measurements described favorites honestly. Both pick rules built on them were not supported — one on every gate, the other on the two gates that distinguish a real effect from a lucky price pattern. Whenever a measurement is presented to you as a method, ask which of the four claims is being made, and ask for the rule, the baseline, the interval and the duel.

Third: use descriptions to set expectations, prices to make bets, and passes to protect both. A description can tell you which favorites to respect, which to doubt, and which races to leave alone. It cannot tell you who wins instead. The laboratory needed a full population of races to learn that its two most truthful measurements could not do that. You can have the lesson for the price of this chapter: believe what the count says about the favorite, and do not ask it to say more.

Chapter 15 — Committees of numbers

Ask a room of handicappers who wins a race and something reassuring happens: they mostly agree. Ask five handicapping methods, or five experts, or five computer systems, and the same reassurance arrives in a more official costume — a consensus. The favorite is the favorite on most sheets. The second choice is the second choice. Where everyone agrees, the pick feels less like an opinion and more like a measurement, the way five thermometers reading the same temperature settle the question of the weather.

This chapter is about why that feeling is worth almost nothing on its own, and about the three separate scoreboards a committee has to be graded on before any of its numbers mean anything. The laboratory runs a committee of its own — its daily selection is made by a voting method, not by a single ranking — and it also, on one November day, invited a committee of outside systems to file picks on the same card. Both committees produced an agreement number that looked, at first glance, like evidence. Both are in this chapter, graded properly. The grading is the chapter.

The three scoreboards

A committee — any panel, voting system, or consensus method — can be scored three ways, and the scores do not substitute for each other.

Agreement asks: how often does the committee land on the same horse as some reference pick — the sheet's own selection, the favorite, another panel? Agreement is a comparison

between two opinions. No race has to be run to compute it. It can be tallied the moment the picks are filed, which is exactly why it is the scoreboard everyone reaches for first: it is available immediately, it is always a clean fraction, and it feels like corroboration.

Accuracy asks: how often do the committee's picks win? This scoreboard needs the races run and counted, all of them, with the denominator stated. It is slower and less flattering than agreement, because reality keeps its own tally and does not care how the votes went.

Financial performance asks: what did the picks return on win bets, at the prices actually paid? This is the slowest scoreboard and the rudest. A committee can agree often and win often and still return less than the takeout takes, because winning at short prices is a way to lose money in good company — Chapter 1's lesson, wearing a committee's clothes.

The discipline of this chapter — and it is the same four-claims discipline the whole book runs on — is that a number from one scoreboard may never be spent on another. Agreement is not accuracy. Accuracy is not a return. A committee that shows you only its agreement number has shown you the one score that costs nothing to produce and proves nothing about the races. Keep the three scoreboards separate and a surprising amount of handicapping marketing becomes unreadable, which is another way of saying it becomes readable correctly for the first time.

The laboratory's own committee, on all three

Start at home. The laboratory's selection method is a committee in miniature. It takes the morning line's top four horses — the crowd's own short list — and scores them on five factors: the morning line itself, the laboratory's handicap weight, its own speed figure, the early pace percentile, and a draw index read from a frozen table. Each factor votes by points, first to third. The sheet's own filed pick, made by a separate rule, sits in the vote as an additional column carrying its own points. The horse with the most points inside the four becomes the selection, subject to a price cap that can hand the selection to the favorite when the committee's choice is priced too long. Five voters, a guest column, a cap. Nobody would call it simple, and Chapter 10 has already billed it for its complexity. Here it is simply graded.

Its population, stated with the care this book requires: the lab's own settled product record: the filed cards of November 2–4, 2026 — 111 races. That is a small sample, three days, and this chapter says so plainly wherever these figures print. Of those 111 races, 95 carried a filed selection.

Agreement. The consensus step changed the official pick in 12 of 95 races. In other words, the pick and the consensus agreed in 83 of 95 races. Agreement between a voting method and the pick it votes alongside sounds like harmony, and the interesting part is the minority where they part. Agreement by itself certifies nothing. Only the races run afterward can grade either horse.

Accuracy. On the lab's own settled product record: the filed cards of November 2–4, 2026 — 111 races, the §26 selection book ($2 win on the consensus selection; a scratched selection counts as a refund) covers 95 races with a filed selection. It placed 90 bets and had 25 winners, a win rate of 27.8%. The official-pick book (one official leg per race, backup chain applied) covers 95 legs. It had 27 winners, a win rate of 28.4%.

Hold those against the paired favorite reference. On the 85 races where both the selection and the favorite started, the selection had 24 winners, a win rate of 28.2%, and the favorite had 25 winners, a win rate of 29.4%. Read that without flinching, because it is the chapter in one line: on this record, the selection did not beat the favorite.

Financial performance. On the same rebuilt record, the §26 selection book staked $180.00 and returned $101.20, for an ROI of −43.78%. The official-pick book staked $190.00 and returned $118.20, for an ROI of −37.79%. On the paired 85 races, the selection returned an ROI of −41.76% and the favorite returned an ROI of −38.88%. The paired difference was −2.88 points. The selection did not beat the favorite on this record — in accuracy, in return, or in the paired difference. That is a three-day record. It is small and swingy. It does not prove a long-run rate. It is, however, the lab's publishable record, and this book prints it at its actual size instead of borrowing a larger population it may not publish.

The laboratory also ran a separate economics test on an earlier form of this committee idea — a different pick object, and the chapter keeps it labeled as one. Its corpus numbers are struck in this rebuild. Its narrative and its gate verdict stand: the paired edge was concentrated in one band of favorite prices and in one

calendar half, and its interval crossed zero. The gate verdict was the label Chapter 1 introduced: ANNOTATION ONLY — a pick rule with evidence behind it and no proven wagering meaning. No control-population numbers are printed in this rebuild.

One more committee fact belongs here, because casual players over-read it in both directions. The committee only votes inside the morning line's top four. On the lab's own settled product record: the filed cards of November 2–4, 2026 — 111 races, the winner sat in the morning-line top four in 84 of the 111 card races, or 75.7%. The morning-line favorite over all 111 card races, started in 101 races, had 27 winners, a win rate of 26.7%; $202.00 staked returned $111.10, for an ROI of –45.00%. As a feat of narrowing, the top-four fact is real: in three races in four, on this small record, the winner stood in the crowd's own short list. As a feat of picking, it says almost nothing — the committee's job begins where that fact ends, choosing one of the four, and in the remaining races the winner was never on its ballot at all. A short list is not a selection. Agreement inside a committee about which four horses matter is the cheapest agreement of all, because the morning line did the work before the committee sat down.

So the laboratory's own committee finishes with three different grades on three scoreboards, on a small three-day record: it agreed with its own sheet in most races; it named winners at about the same low rate as the other books on the same cards and did not beat the favorite on the paired races; and it returned a clear loss. Any single one of those sentences, quoted alone, would mislead you about the other two. That is not a defect of

the committee. That is what honest grading looks like, and it is the standard the rest of the chapter applies to the visiting committee.

The visiting committee

On November 5, 2026, the laboratory ran an experiment of a kind that has become common and is almost never graded. Six outside AI systems — this book names no models and no makers, here or anywhere — were given the day's published card and asked to file their own selections. The exercise had the shape of an audit: independent systems, the same races, their picks set against the sheet's. What came back is a perfect specimen of the agreement scoreboard, and this chapter presents it exactly as it stands in the lab's ledger, including the parts that do not flatter the exercise.

The fullest run filed 41 listings: the card's 37 pick races, plus four display-only races where it also took the consensus leader. On the 37 pick races it agreed with the site's filed selection 31 times. It filed five outright disagreements, and one further pick that was not a disagreement at all — the site's selection in that race was confirmed scratched that morning, and the outside system, reading the same changes reports, named a different horse in the vacated race. Thirty-one of 37. As a fraction, 83.8%. It is the kind of number that gets quoted, and it was, in the day's own internal talk, the number that made the exercise feel successful.

Now grade the rest of the panel, because the panel was six seats and the number above is one seat. A second system agreed 17 times in 17 — a spotless record that covered only the first 17

pick races of the day before it stopped, and whose own closing note said it had retained the morning-line leader where its data was thin; its 17 picks were identical to the booked selections, which means any later score for it would have equaled the card's own score on those races by construction, not by judgment. A third system returned only three races: one verified agreement and two verified disagreements — a 33% agreement rate on a sample too small to grade, which did not stop it being a third of a "panel." A fourth filed six picks at a single track: two agreements, three disagreements, and one pick in a race where the site had filed no selection at all — and it filed them from outside tip sheets, because the site itself would not load for it; it agreed and disagreed with a sheet it had never seen, through a summary of other people's opinions. Two seats filed nothing: one could not load the site on two attempts and said so plainly, and one hit a usage limit. Six seats: one full card, one half card, one three-race fragment, one track's worth of secondhand picks, and two empty chairs.

And the fullest run itself carries a condition that the agreement number does not mention. Before filing, it had read the site's picks sheet — all 37 filed races — and the consensus document besides. Its selections were made with the sheet's answers open in front of it. That does not make its 31 agreements worthless; a reader can always disagree, and it did, six times. But it changes what the number measures. A system that reads the filed selections and then files its own is partly grading the sheet and partly transcribing it, and an agreement rate computed on transcription-plus-judgment cannot be spent as independent corroboration. Independence is not a vibe. It is a condition of the test, and the condition was not met.

The six disagreements, and the one useful thing

The day was not wasted, and the reason is instructive: the value came from the disagreements, not the agreements, and its value was editorial rather than statistical.

The fullest run's six differing picks were adjudicated before the races, against the filed sheet data, by the laboratory itself. The sheet's reading was upheld in four of the six — once decisively, once by a narrow margin, once outright, and once on points in the race where the outside case was called the strongest of the six. Two were left genuinely open: one was a dispute not with the race reading but with the price cap that had replaced the committee's choice with the favorite, and one was a race with four Unknown starters, where the laboratory rated its own confidence low either way. The pattern across the six was the pattern Chapter 13 measured at full scale: five of the six differing picks took the horse with the loudest single factor — the top early figure, the top weight, the top speed figure, the morning-line favorite — against a filed selection whose support was spread across several factors at once. Loud single factors are how plausible disagreements are manufactured. The count in Chapter 13 is what they are worth as picks: as a group, horses carried by one top factor win rarely and return badly.

The run also did one operational thing well. Its morning report flagged scratches and track-condition changes across the card, and every one of those reports was checked that morning against the official changes page and confirmed. That is a real contribution, and notice its category: it is fact-checking, not handicapping. It made the day's record more accurate. It said nothing about whose picks were better. The laboratory folded the lesson into its own routine — a standing morning changes

check now covers the same ground — which is the correct fate for a useful clerical finding: adopt the procedure, grade the picks separately.

The scoreboard that was never filled in

Here is the fact this chapter exists around. The outside panel's picks were not scored. The day's ledger file was opened with standing text that describes outside picks being priced as two-dollar win legs at the evening build, and the day's entry was marked pending settlement — but the day's own record states that scoring was waived by the laboratory's owner, in plain words: no need to score them. The selections remain on file, exactly as filed. No win rate for the panel exists. No return for the panel exists. There is no accuracy scoreboard and no financial scoreboard for the visiting committee, and under the waiver there never will be for that day; the races are run, and a score computed now would be computed under a decision made after the fact, which is a different and lesser kind of number.

So the episode's entire evidentiary estate is one number: 31 agreements in 37 races, filed by the one full run, under non-blind conditions, alongside a 17-for-17 partial run that agreed by construction and a scatter of fragments. Set that beside the laboratory's own committee, which needed a large count to earn the right to make a modest claim and still had its economics ruled unproven. The panel's number is not smaller than the laboratory's numbers. It is a different *kind* of thing. It counts resemblance between two sets of opinions, one of which was written with the other in view. Resemblance feels like evidence because in ordinary life it often is: when two people independently describe the same stranger, the description gains

credit. The credit comes entirely from the independence. Remove it — let the second witness read the first one's statement, let half the witnesses leave early, let two of them never arrive — and the resemblance that remains measures something much plainer: that similar inputs, read by similar systems, produce similar picks. Nobody needed 37 races to learn that.

It is worth saying precisely what would have made the panel evidence, because the gap between the exercise as run and the exercise as evidence is the whole lesson. Decide in advance that the picks will be scored — as win legs, against the settled results, all of them, including the fragments and the empty chairs, which count as the panel's coverage and not as races that never happened. Keep the panel blind to the filed selections, or drop the word independent and grade the exercise as review instead. Run it for a season, not a day, because 37 races cannot separate systems whose true accuracies differ by the few points that matter — the laboratory's own committee and the plain favorite sit close together and need many races to be told apart with any confidence, and even then a money difference can stay inside an interval that includes zero. Then publish all three scoreboards side by side: agreement, accuracy, return. A panel graded that way could teach the laboratory something it does not know. A panel graded on agreement alone teaches it something it already knew and flatters it in the bargain.

And notice, finally, the direction of the flattery, because it is the surest sign the number is doing no work. Had the fullest run agreed 31 times with a sheet whose selections, on the lab's own settled product record: the filed cards of November 2–4, 2026 — 111 races, returned −43.78%, the agreement would certify that

the visitor resembles a money-losing method. Had it disagreed 31 times, the disagreement would have been quoted as fresh eyes finding the sheet's errors. The same number, read both ways, always in the quoter's favor. A measurement that confirms every conclusion is not a measurement. It is a mirror.

The three lessons of committees

First: keep three scoreboards, and never let one pay another's bills. Agreement is computed before a race is run and proves resemblance. Accuracy needs every race counted. Returns need the prices. On the lab's own settled product record: the filed cards of November 2–4, 2026 — 111 races, the laboratory's own committee agreed with its sheet in 83 of 95 races, named winners at 27.8% — and did not beat the favorite on the paired races — and returned −43.78%, with its economics formally unproven and ruled ANNOTATION ONLY. Those are three grades, not one, and the committee is all of them at once.

Second: agreement is only as good as the independence behind it. A panel that reads the filed picks before filing its own has told you that similar readers reach similar readings. Partial runs, secondhand picks and empty chairs are part of the panel's record, not footnotes to it. Count the coverage with the agreements, or do not quote either.

Third: an unscored panel is a mirror, not a judge. The visiting committee's 31 of 37 was the whole of its evidence because its scoring was waived, and a waived scoreboard cannot be filled in later from memory or mood. If a committee's opinion is worth soliciting, its picks are worth scoring — in advance, in full, and

on the same three scoreboards you apply to your own. Anything less is asking the room to applaud, and the room, given the sheet in front of it, usually will.

Chapter 16 — Track bias without superstition

Every track you visit has a story attached to it, and the story is usually told as a personality. The rail is golden there. Closers are dead there. Speed holds there in the fall. The story gets repeated in the grandstand, in the program notes, and in the comments under last week's results, and it has one property that should make you suspicious the moment you hear it: it never expires. A personality is permanent. It explains last April and it will explain next October, and because it explains everything, it can never be wrong. A race that fits the story confirms it. A race that does not fit gets a excuse — the pace was wrong, the rail was scraped, the winner was simply better — and the story walks on untouched.

This chapter argues for a smaller, duller, and more useful idea. A track does not have a personality. A track has a surface, and a surface has a condition. Rain puts water in it. A dry week and a harrow take water out and move the cushion around. Wind dries one side faster than the other. A turf course firms up, or softens, or gets worn along the inside paths where every horse has run for three weeks. A maintenance crew can change the surface between Tuesday and Saturday without asking anyone's permission, and a change of weather can change it between the third race and the eighth. What players call bias is the fingerprint of those facts on recent results: for a short stretch of days, winners kept coming from one part of the field — the early leaders, the stalkers sitting just behind them, the inside posts, the outside posts — more often than chance and field composition would lead you to expect.

Notice what that definition does. It gives bias a memory, and the memory is short. It ties bias to a surface and a distance, not to a track's reputation, because a wet sprint strip and a firm turf route course at the same address are two different pieces of ground. And it makes bias a claim you can check: count the recent winners, sort them by how they ran, and see whether one group is carrying a share large enough to matter. If the count does not show it, the story does not get to say it anyway.

That is the whole method. The rest of this chapter is one worked count, done in full, and the rules the lab uses to keep itself honest while doing it.

The worked count: Keeneland at the close of its fall meet

In the first days of November 2026, the lab ran a bias study on Keeneland, and it is a good specimen because the calendar handed us a natural experiment. Keeneland's regular fall meet ran from October 2 through October 24. The track then went dark for five days, October 25 through 29, and closed its season with two championship cards: the Breeders' Cup, on Friday October 30 and Saturday October 31. So the recent record holds two kinds of racing on the same ground within the same two weeks — ordinary meet racing, and two days of large-field stakes racing that draw horses, riders, and running styles from everywhere. If bias were a personality, both populations should show the same fingerprint. If bias is a short-window condition, they might not.

The method was plain counting, and the chapter states it in full because a bias claim you cannot recount is a story. The source was the Equibase full-card chart for each raceday — the settlement source of record, the same charts the lab settles its own cards from. Every winner was sorted into a bucket by surface (dirt or turf) and distance class (sprints under eight furlongs, routes at eight furlongs and up). Each winner was then labeled by the trip it actually ran, read from the chart at the first call: a winner that led, or ran second within one length, was Early; a winner within three and a half lengths without leading was a Stalker; a winner more than three and a half lengths back was a Closer. Those labels describe trips, not reputations — a horse the past performances call a closer can win with a stalking trip, and in this count it lands in Stalker, because the count is about how races were won, not about what the winners usually do.

Two discipline rules governed what the count was allowed to say. First, a lean prints only when one style wins at least 40% of a bucket and clears the next style by at least eight points. Anything weaker is reported as a split, in those words. Second, a bucket with fewer than eight styled winners is labeled thin and states no lean at all — it shows its counts and stays silent. Of the 134 winners in the windows studied, 127 carried a trip that parsed and entered the style shares; the other seven kept their identity and post group from the chart's mutuel table but stayed out of the style counts, and the study names their races rather than quietly dropping them.

Hold those two rules in mind. They are the difference between this chapter and the grandstand story. The story states a lean on any evidence, including one afternoon. The count states a lean only when the evidence clears a bar set in advance, and it is willing — often — to say "split" or "thin" and stop.

What fourteen days showed

The verdict window was the last fourteen calendar days, October 18 through 31: seven racedays — five regular days on October 18, 21, 22, 23, and 24, plus the two Breeders' Cup days — and 68 races.

On dirt, the read was steady and strong. Of 49 dirt races, 48 winners carried a parsed trip. Stalkers won 62% of them. Early winners took 35%, and deep closers took 2%. Wire-to-wire winners — horses that led at the first call and never gave it up — accounted for 23%. Split the dirt by distance and the same shape held in both halves: in 25 sprints, stalkers won 60% and early horses 36%; in 24 routes, stalkers won 65% and early horses 35%, with no closer winner at all among the 23 styled route winners. By the lab's rule — at least 40%, clear by at least eight — dirt leaned Stalker, plainly.

On turf, the same fourteen days showed no lean at all. Of 19 turf races, 16 winners carried a parsed trip, and they divided exactly: early winners 50%, stalkers 50%, closers 0%. The honest line for that bucket is the one the machinery prints when the bar is not cleared: wins split. Underneath the split, the distance split added a detail worth noticing precisely because the chapter will not build on it: in 15 turf routes, the 14 styled winners went 57% to stalkers, a Stalker lean inside that one band, while the four

turf sprints produced only two styled winners — a thin cell that states counts and no lean. A split at the top level and a lean in one band can both be true. They answer different questions, and confusing them is how a split becomes a story by dinnertime.

That is the fourteen-day picture: dirt leans one way with room to spare; turf declines to lean. Now take the two championship days out and count the same fourteen days again — the five regular racedays only, 46 races.

Dirt does not move. Without the Breeders' Cup days, 34 styled dirt winners went 59% to stalkers, 38% to early horses, 3% to closers, and wire-to-wire winners rose to 29%. Stalker at 59% with the championship days removed, Stalker at 62% with them in. Two different populations of horses, the same fingerprint on the same surface. That stability is what a real short-window condition looks like, and it is worth more than the size of either number: a lean that survives removing the two loudest days of the meet is describing the ground, not the guest list.

Turf moves completely. Without the Breeders' Cup days, the ten styled turf winners went 60% to early horses and 40% to stalkers, with wire-to-wire winners at 40% — an Early lean under the lab's rule. Put the championship days back in and the lean dissolves into the 50–50 split you saw above. The Breeders' Cup turf winners, in other words, came from stalking trips in large fields, and there were enough of them to erase the regular meet's early-speed read inside a fourteen-day window. Neither version is the track's personality. The regular days describe how Keeneland turf played for ordinary fields in late October. The championship days describe how championship turf races, with their larger fields and their deeper benches of late runners, were won on that ground. A player who carried the regular meet's

"turf is early" story into Breeders' Cup weekend would have been betting a true statement about last week against a different population of races. A player who carried the weekend's "turf winners stalk" read into a ordinary Thursday — if the meet had continued — would have made the mirror-image mistake.

And on the championship dirt, one number deserves its own sentence because it shows how far a subpopulation can drift from the window that contains it. Across the two Breeders' Cup days alone — 22 races, the only racing in the last-seven-days window, since the track was dark from October 25 to 29 — dirt winners went 71% to stalkers among 14 styled winners, and exactly one of those 14 winners led at the first call: a wire-to-wire rate of 7%, against 29% on the regular-meet dirt inside the same fourteen days. Read that window for what it is. It is a championship-card profile, not a meet bias: the fields averaged about ten horses against about 8.6 on the regular window days, nearly every race was a stakes, and six of the eight turf races in it were routes. The seven-day window did not malfunction. It answered a different question than the one a casual player usually asks, because the week it covered contained no ordinary racing at all.

The shorter window, and the single day

Cut the window to the last three racedays — October 24 plus the two championship days, 32 races — and the count still leans Stalker on both surfaces: dirt at 68% among 22 styled winners, turf at 75% among only eight styled winners, a cell small enough that the chapter reports it and leans on it lightly. Cut it differently — the last three regular racedays only, October 22, 23, and 24 — and dirt still leans Stalker at 59% on 22 styled

winners, while the turf cell shrinks to five styled winners, prints Stalker at 60%, and is labeled thin: counts, no lean. Inside that regular-only cut, dirt routes split evenly, early 50% and stalkers 50%, against the 64% Stalker read the same band shows once the championship days join the three-day window. Same track, same week, three defensible windows, three slightly different answers — one of them a lean, one of them a split, one of them too thin to speak. That is not a defect in the method. That is the method telling you the truth about sample size: as the window narrows, each winner moves the percentages more, and the honest output degrades gracefully from a lean to a split to silence.

The single day degrades all the way. Inside the fourteen-day window, no individual raceday had enough styled winners in any bucket to clear the eight-winner floor, and the day-by-day counts swing the way small counts swing. On October 18, the six styled dirt winners went 67% to early horses. On October 23, the seven styled dirt winners went 0% to early horses and 86% to stalkers. On October 22, early horses took 71% of seven styled dirt winners; five days of racing later, on the first championship day, the six styled dirt winners split 50–50, and on the second, stalkers took 88% of eight. A player reading bias off single days would have held four contradictory opinions in nine days, each one supported by that day's results, and would have bet each one with the confidence of a man who had "seen it himself." He had seen it. Seeing a day is not the problem. The problem is what a day can support: eight to twelve races, split by surface and distance into cells of two to eight winners, where one horse changing its trip moves a share by twelve points or more. The

count's thin rule exists for exactly this man. It is not there to hide information. The counts print. It is there to stop a count of six from wearing the uniform of a finding.

The meet as a whole is not a bias

One more table from the same study, because it draws the chapter's central line. The lab also counted the regular fall meet's charted days as a whole — twelve racedays from October 9 through 24, 112 races. Across the meet, dirt went to stalkers at 64% among 78 styled winners, with early horses at 35% and closers at 1%, and wire-to-wire winners at 24%. Turf across the meet went to stalkers at 55% among 29 styled winners, early horses at 45%, no closer winners. If you stopped there, you would write the sentence the grandstand writes: Keeneland is a stalker's track. The study refuses to call that table a bias, and labels it a baseline — a description of the track's character across the meet, kept for contrast. The reason is in the post positions. Across the full meet, winners from posts 7 and out took 43% of dirt races, the largest share of any post group. In the late windows, the tilt reversed: posts 1 through 3 led the dirt counts, taking 41% in the fourteen-day window, 43% on the championship days, and 50% in the three-raceday cut. A season-long average buries that reversal. It hands you one number for the meet that describes neither the meet's early weeks nor its close, and it invites you to bet the buried average on the next card as if the ground had signed a contract. The ground signs nothing. The baseline tells you what kind of track this generally is. The short window tells you how it has played lately. Only the second one is a bias read, and it expires.

That expiry is the point the superstition cannot absorb. A personality is a claim with no date on it. A bias read is a claim with nothing but dates on it: these racedays, this surface, this distance class, this many winners, this share. When the dates stop being recent, the claim stops being a bias claim and becomes history. Keeneland's season ended on October 31, and the study says so in plain words: these numbers describe how the track played at the close, and they carry no forward window until racing resumes. The most honest bias line in this chapter is a line that has already expired.

How the lab reads bias every day

The Keeneland study was a hand-built count for one track at one moment. The daily card needs the same reading for every track on the card, every morning, and the machinery that does it — a program called fetch_bias — is built to enforce this chapter's discipline when nobody is watching. It works in buckets, always: one bucket for each combination of track, surface, and distance class on the card, sprints separated from routes, dirt from turf from synthetic. A card never gets one bias line for a whole track. It gets a line per bucket, or it gets an honest no-read line for that bucket.

For each bucket, the machinery computes four layers of evidence and displays the first available one in a fixed priority order. First, the lab's own settled results from the last seven days — the freshest layer, built from the charts the lab itself settles each night, counting the winner's running style and post group. A seven-day bucket is small by construction, so its floor is eight races; the lean gates still apply in full. Second, a published track-trends source over the last fourteen days, with a floor of

fifteen races. Third, the lab's own settled results over the last thirty days, floor fifteen. Fourth, the published source over thirty days, widened once to sixty when the thirty-day sample is thin. The first layer that clears its floor supplies the line, and the line carries its source and its race count in its text, so a reader can see not only what the read is but whose count it came from and how much racing stands behind it.

Two more rules do the honesty work. The lean rule is the same one the Keeneland study used: a style lean prints only when the leading style holds at least 40% and clears the next style by at least eight points. Below that bar, the line says the wins are split and lists the shares. And when the two freshest layers — the lab's own seven days and the published fourteen — both state a lean and the leans conflict, the machinery does not average them into a compromise lean that neither source holds. It prints a split, and it names both reads: this share in our settled results, that share in the published read. A disagreement between two honest counts is information. Blending it away would manufacture a third opinion that no one counted. The lab labels disagreements instead of blending them, on the card, in the open.

And the last rule is the one that keeps bias in its place in this book's system: the bias line is display only. It prints beside the race. It never reorders a ranking, never moves a pick, never enters a probability. That restraint is not modesty. It follows from what kind of claim a bias line is, which is the subject the book has returned to in every part, and which this chapter now applies to bias directly.

Four claims, and which one a bias line makes

A bias count, honestly done, supports an association claim: over these recent races, winners came from this part of the field this often. That is all. It does not, by itself, support a ranking claim — it does not follow that the stalker in today's fifth race should be ranked above the early horse beside it, because the count describes winners as a group and says nothing about which of today's horses runs which trip, or at what price. It does not support a probability claim: "stalkers won 62%" is a share of past winners, not a chance for any horse in front of you, and treating a winner-share as a probability double-counts the base rates of how many stalkers start each race in the first place — a style that supplies half the starters and half the winners is showing you nothing. And it does not support a returns claim. Nobody in the sources for this chapter bet the Keeneland leans and counted the money. A lean can be real as a description and worthless as a bet, if the crowd has seen the same recent winners you have — and the crowd has, because they were run in public — and has already shortened the prices of every horse that fits the pattern.

What would promote a bias read toward those stronger claims? The same ladder the rest of the book uses. A rule written in advance — bet this kind of horse, in this kind of bucket, when this lean prints — counted over every race it touches, compared against simply betting the favorite in the same races, with the money counted at the end. Until a bias rule climbs that ladder, the lab treats its own bias lines the way it treats every other unpromoted description in this book: as context that can lower your confidence, shape your reading of how a race may run, and

give you one more reason to leave a race alone — the subject of the next chapter — but never as a reason to promote a horse the rest of your reading does not support.

Here the chapter owes you the same disclosure Chapter 1 gave you about the lab's selections, applied to the lab's own window. The seven-to-fourteen-day window this machinery uses — seven days for the lab's own freshest layer, fourteen for the published recent layer — is the lab's working practice. It is not a validated law, and the lab has not run the test that would make it one: no count in the sources behind this chapter compares window lengths against each other on future races and crowns a winner. The window rests on reasoning, not on a result. Surfaces change on the scale of days, so a window of months buries the current condition under old ground, as the Keeneland baseline buried the late inside tilt. Weather and maintenance operate inside a week or two, so a window of a day or two collapses into the noise you saw in the day-by-day counts. Seven to fourteen days is the working compromise between burial and noise. It may be right. It has the same status as a handicapper's rule of thumb that has not yet been counted, and this book's standard requires the chapter to say so rather than dress a practice up as a finding. If the lab ever freezes a set of candidate windows, runs them forward, and counts the outcomes, the window will earn a stronger label. Until then: practice, pending formal validation, not a universal law — and certainly not a law of other tracks, other seasons, or other surfaces, each of which gets read in its own buckets or not at all.

Reading a bias line like a casual player

You do not need the machinery to use the discipline. When any source — a track program, a tip sheet, a broadcast, a friend — hands you a bias claim, ask it the questions this chapter asked.

Ask for the window, and distrust any answer that is a season, a meet, or a reputation. "Lately" should mean named dates inside roughly the last two weeks, and the claim should die when those dates stop being recent. Ask for the bucket. A claim about a track that does not separate dirt from turf and sprints from routes is averaging pieces of ground that play differently — at Keeneland in the same fourteen days, dirt leaned Stalker at 62% while turf split 50–50, and a single "Keeneland favors stalkers" sentence would have been wrong about the turf and sloppy about the dirt. Ask for the count behind the share. Sixty-two percent of 48 styled winners is a statement. Sixty percent of five is a coin that landed kindly, and the honest source will call it thin or stay silent; if yours calls it a trend, find a better source. Ask what the claim excludes: the Keeneland turf read changed because two championship days entered the window, and a source that never separates unusual cards — championship days, festival weeks, the sloppy afternoon when half the card moved off the turf — will hand you blended numbers for populations you are not betting. And ask, finally, what the claim is for. If the answer is a story about the track's character, thank it and move on. If the answer is a description of recent winners, use it as one: let it temper a bet on a deep closer on a strip where deep closers took 2% of dirt winners in the last fourteen days, let it make you look twice at a stalker you already liked,

and let it add a reason — never the only reason — to pass a race whose shape runs against everything the recent ground has rewarded.

What you should not do is the move the superstition exists to make: turn the description into a destiny. Horses do not know the bias line printed beside their race. An early horse with a clear lead and no pressure can win on a stalker's week, and at Keeneland during the championship days one dirt winner in fourteen did lead at the first call — few, but not none, and that one paid somebody. The bias read tells you how the recent races were won. The race in front of you will be won how it is won. The distance between those two sentences is where casual players either keep their money or donate it to a story.

The three lessons of the bias count

First: bias is a condition, not a character. It belongs to a surface, a distance class, and a short run of recent dates. A claim without dates, a surface, and a count is a superstition wearing a statistic's clothes.

Second: let the count refuse to speak. A lean needs a bar — this lab's is 40% of the bucket, clear by eight points, on at least eight winners for the freshest layer and fifteen beyond it — and below the bar the honest words are "split" and "thin." A source that always has a lean always has a story, and a story is what you were trying to escape.

Third: a bias line describes; it does not promote. It is an association claim about recent winners. Ranking a horse, pricing a horse, and betting a horse are three stronger claims,

and each needs its own evidence. Use bias to read a race and, often, to leave one. The next chapter is about leaving races — and about how much that habit, counted honestly, is worth.

Chapter 17 — The races to leave alone

The cheapest decision in handicapping is the one most players refuse to make. It costs nothing to consider, it takes ten seconds, and over a season it is worth more than most systems ever tested in our laboratory. The decision is to bet nothing.

Players resist it for reasons unrelated to arithmetic. They came to play. A passed race feels like a wasted opinion — you did the reading, formed a view, and let the race run without you, and sometimes the horse you liked wins and pays. This chapter asks you to count the pass the way the lab counts everything else: not by the winners you watched someone else collect, but by what the bets you declined would have returned across all the races where you declined them. Counted that way, passing is not the absence of a decision. It is a decision with a measured value, and in the specific race shapes this chapter names, the value is not small.

The lab's own product passes races every day, on purpose, under written rules. That surprises people who meet the product first as a sheet full of picks. The sheet is full of picks because it covers many races. Underneath it sits a protocol — a frozen, numbered rulebook, currently Selection Protocol v3.17 — and several of its rules exist for one purpose: to name the races and the horses the operation will not file a pick on, no matter how interesting they look. This chapter walks through those rules in plain language, adds one measured warning sign that falls short of a rule and must be labeled as such, and then works three pass

decisions the way you would meet them on a Saturday. It closes with the season arithmetic: what a habit of passing the named shapes costs, and what it saves.

Keep the book's standing discipline in view as you read. Each item below makes a different kind of claim. A filing rule is a decision the lab adopted after a test; the test figures behind it are returns claims about a defined population, and the chapter names the population every time. A diagnostic screen is an association claim: it describes a group of horses honestly and has not been adopted as a rule, and the chapter will not let it borrow a rule's authority. And the worked decisions at the end are illustrations built from the documented situations — they show the reasoning applied to a race shape, they are not results, and no figure in them is a measurement except the figures carried in from the tests above.

No data is not neutral data: the Unknown in the field

Start with the horse you know nothing about. Every card has them: a starter whose record gives the lab no usable pace line — no prior race it can score, no figure it can stand on. The sheet marks such a horse Unknown, lands it at the field average for display, and keeps the race. The old instinct, in this lab and everywhere else, was one of two errors. Either you throw the race out entirely, punishing nine horses you can read for one you cannot, or you treat the Unknown as an average horse and handicap the race as if the gap in your knowledge were a fact about the horse. The protocol does neither, and the reason matters for your own betting, not just for the sheet.

An average is a measurement of horses that were measured. The Unknown was not measured. When the lab's probability model — the blend, the part of the system that turns a field's figures into calibrated chances — meets an Unknown, the protocol's statement is exact: the horse has no blend probability. Not zero. Not the field average. None. Zero would be a claim that the horse cannot win, and nobody knows that; a first-time starter from a sharp barn wins races every week, and a shipper with no usable line in your records may be the class of the field in everyone else's. The field average would be a claim that the horse is typical, and nobody knows that either. The honest value for a chance you cannot estimate is not a polite number. It is a blank, and the protocol keeps it blank: the horse cannot trigger the model's deviation rules, cannot satisfy them, and a race containing an Unknown is not counted as covered by the model at all. The race files under a flagged fallback, with the gap declared, or — for you, at your own window — it does not file at all.

Two display details complete the rule, and both exist to stop a blank from masquerading as a number. The field-average landing is a graph convention, a way to draw the horse on a pace picture without inventing a shape for it; it is never fed to the model as a probability, and this book's earlier chapters apply the same rule to you: a bar drawn at the average is a placeholder, not a reading. And a horse with no line in the store shows a labeled par weight, marked as par, display only — never a signal, never a reason to move it up or down. What the sheet will still do is rank the horse by its morning line, because the anchor of the whole system is the market's order, and the market has seen things your pace records have not: workouts, entries, the judgment of the people who set the line and the

people who will bet against it. Respect the ranking. Question everything else you think you know about the race — including whether you know enough to bet it.

That last step is yours, and here is how the lab's practice translates to a casual player's rule. One Unknown in a field you otherwise read well is a declared gap: you can bet the race if your reason never touches the Unknown — but be honest that your reason rarely survives contact with it. If your pick's case is "he controls the pace," and the Unknown is an unraced speed-bred sprinter drawn inside him, your case has a hole in it shaped exactly like the horse you cannot read. If the Unknown is the favorite, the situation is sharper: the crowd's top choice is a horse your records cannot score, which means the price structure of the whole race rests on information you do not hold. And two or more Unknowns in one field is not a race with gaps. It is a gap with a race around it. The lab keeps such races on the card and flags them. It does not cover them with its model, and it will not make one the day's free race — the free-race rule re-points off any race containing an Unknown, the same way it re-points off an excluded race. For your own money, the translation is short: a race the model refuses to cover is a race you are allowed to refuse to bet. The chapter flags one gap in its own evidence here, because the book's standard requires it: the sources behind this section state the rule and its reasoning, and they contain no win rate and no return figure for races with Unknown starters. The pass rests on a knowledge claim — you cannot price what you cannot read — not on a counted loss. That is a weaker kind of evidence than the next section's, and the chapter labels it so rather than borrowing strength it has not earned.

The turf cells the lab will not pick

The next pass is a counted one. It comes from a test, it survived the test in a narrower form than its authors proposed, and the narrowing is the lesson.

The question was whether the lab's selections did worse on turf than on dirt by enough to justify leaving turf alone. Tested whole — the entire turf surface against the entire dirt surface — the answer was no, or rather, not enough: the turf-versus-dirt gap fell under the test's pre-registered materiality bar, as filed and on the clean universe. Worse, the gap failed the stability check the lab applies before it believes anything: in the first half of the year, turf selections booked about as well as dirt, and the whole drag was a second-half phenomenon. A surface-wide turf ban would have been a permanent rule built on half a year. The lab refused it, and whole-surface turf exclusion sits on this book's closed list: tested, not adopted, not to be revived without new evidence.

What the same test supported, applied cell by cell under the same rule, was narrower and uglier. Two turf cells lost in both halves of the year, by margins that cleared the bar. Non-maiden turf races — turf races for horses that have already won, the allowances, claimers, and stakes that make up most of a turf card — returned clearly less than the dirt book, worse than dirt in both halves, with a losing duel record against the favorite baseline. Turf races with filed fields of nine or more starters also returned clearly less, worse in both halves, with a losing duel record. And one turf cell pointed the other way: turf maidens returned better than the dirt book, on the clean universe and as filed. Maidens, remember, are the least-known

fields in the game — and on turf they were the lab's least-bad turf population, which is a useful reminder that "hard to know" and "bad to bet" are different properties, measured separately.

The adopted rule, §25 of the protocol, files no pick when either losing cell matches, judged on the race as filed at filing time. Cell A: the surface is turf and the race is not a maiden. Cell B: the surface is turf and the filed field has nine or more starters. No pick means the full list: no paired selection, no single, no counter leg of any kind, and no eligibility for the featured races or the free race. The race still appears on the card and the picks sheet with its pace graphs and all its display data — the exclusion is from selections, not from coverage, because the reader may still want to watch it, and because a rule that hides its own exclusions cannot be audited. Turf maidens and turf fields of eight or fewer file normally. The field count is the field as filed, before race-day scratches: a race filed with nine that scratches down to eight still files no pick, because the count operable at pick time is the only honest count a filing rule can use. And a surface change re-judges the race: a race moved off the turf comes out of the cells and files as a normal race would, and a race re-judged into a cell has its filing withdrawn, not redone.

For your own card, the translation needs no protocol. If you bet one horse per race at the kind of prices the lab files, the two named turf shapes have a counted cost: across a large count of selections they returned materially less per dollar than break-even — worse than the favorite baseline this book built Chapter 1 on, worse than the lab's book as a whole. You are free to believe your turf-route reading beats the lab's. The burden the book sets is the burden of Chapter 4: write your rule down,

count every race it touches, compare it to the favorite on the same races, and let it die if it loses. Until your count says otherwise, the counted answer for non-maiden turf races and big turf fields is that they are races to watch.

One correction belongs in this section, in the open, because a draft brief for this chapter described a further no-file list — synthetic surfaces and named tracks, filed under the protocol's §17 — and the correction is the chapter's own standard applied to the chapter. Read the current protocol: no such list exists in Selection Protocol v3.17. The §17 that once stood in the rulebook was removed in version 3.3, and the removal stands in the current text; §25 itself states, in terms, that dirt and synthetic are untouched by the turf exclusions. The lab's current product files picks on synthetic surfaces, and this chapter will not tell you it declines races it demonstrably files. If a future version adopts a synthetic or track exclusion, it will arrive the way §25 arrived — a pre-registered test, a numbered version, a record — and a future edition of this chapter will cite it. Until then, the no-file list in force is the short one in this chapter: the two turf cells, the price ceiling below, and the coverage refusals around the Unknown. A short list you can verify beats a long one you cannot.

The price ceiling: no leg over 8-to-1

The third pass is a horse-shaped one. The protocol's §22 sets a ceiling: no filed leg on a horse whose morning line exceeds 8-to-1. The ceiling covers every kind of filed leg, including the counter-moves and replacement derivations elsewhere in the rulebook. A horse at exactly 8-to-1 files normally — the ceiling removes only prices that exceed it. A horse with no morning

line at all is not affected; a separate proposal to remove no-line horses was tested, failed, and was not adopted, and the protocol keeps the distinction because the evidence drew it.

Read the ceiling correctly, because it is easy to misread as a claim that longshots never win, or that 9-to-1 shots are bad horses. It is neither. It is a filing discipline about where a one-pick-per-race operation's selections actually come from. A system anchored on the market's order, choosing among the market's leading group, does not suddenly find its value in horses the market has priced at the edge of the field; when its derivations reach that far down the page, they are usually reaching — past the evidence, into the part of the field where any story can be told and none can be checked. The ceiling says the reach is not filed. The horse may win. Horses at those prices win every day, at roughly the rate their prices imply, which is the point: at the rate their prices imply, minus the takeout described in Chapter 3, they are a slow way to return less than you staked, and a selection method earns nothing by joining them.

Your version of the ceiling can be stricter or looser than 8-to-1, but have one, and write it down before the card starts. The moment that tests casual players is never the cold longshot in the second race. It is the eighth race, two hours after your third loser, when a 12-to-1 shot's one good figure starts to look like a key that fits the whole day. A ceiling set in the morning answers that moment in advance. A ceiling invented in the moment is not a ceiling. It is a negotiation, and you will lose it.

A warning sign that is not a rule: the horse whose form is on the wrong surface

The last item is the one this chapter must handle most carefully, because it is real, it is measured, and it is not a rule. The lab logs it as a screen — a diagnostic — and the difference between a screen and a rule is the difference this book has policed since Chapter 4. A rule was tested as a way to bet and adopted or refused on its returns. A screen describes a group. It earns a warning label, not a filing.

The screen is called V4, for surface exclusivity, and it flags a favorite whose recent figures were all earned somewhere else. Exactly: the screen flags the favorite when it carries at least two prior lines and every one of its last three prior lines was earned on a different surface family — dirt, turf, synthetic — from today's race. The horse is not switching for the first time; plenty of horses handle a switch. The horse's entire recent evidence, the figures that made it the favorite, was produced on a different kind of ground from the ground it must run on today.

On the primary population, the screen separated the favorite population cleanly. Flagged favorites won less often than unflagged favorites and returned less on win bets. Read both halves of that. As an association, it is a large gap: a favorite whose form is all on the other surface won materially less often than an ordinary favorite. As a returns figure, the flagged group lost more per dollar than the unflagged group — worse, in a population where everything is worse than break-even. And the control agreed: on the control population, flagged favorites again won less often and returned less than unflagged favorites. Two populations, the same separation, in the intended

direction. The lab records V4 as a separating diagnostic screen on both populations — and adopts nothing from it. No pick rule was built on it, no gate was run on it, and the report says in terms that no adoption follows from a screen alone.

So what may you do with it? Exactly what a diagnostic licenses. When the favorite in front of you carries all its recent figures on the other surface, you know one measured thing: favorites shaped like this one, as a group, won less often and returned less than favorites generally, on two large counts. You do not know that this horse will lose — a meaningful share of flagged favorites still won, which is many parades to the winner's circle you will watch if you pass them all. You do not know which other horse beats it; the screen is silent about the rest of the field, and the six-method program this screen came from is a standing warning on that point — again and again in that program, a measurement that described a favorite's weakness honestly still failed, as a pick rule, to name the horse that profited from it. And you do not know that passing flagged favorites, as a betting rule, returns more than betting them, because nobody has run that rule through the gates. What you know justifies a posture, not a system: do not make this favorite the center of your day, do not build exotics that need it, demand a reason beyond its figures — figures earned, remember, on ground it is not standing on — before you bet it at a short price, and count a decision to leave its race alone as a decision the evidence leans toward, at association strength. That is a real use. It is also the whole use. A screen that gets promoted in your head to a law will eventually be bet as a law, at law-sized stakes, on evidence that never carried that weight.

Three worked passes

The situations below are worked examples. They are built from the documented situations above — the Unknowns rule, the §25 cells, the price ceiling, the V4 screen — and they are labeled as illustrations, not results. Any figure in them comes from the tests already cited. The horses are shapes, not animals.

The turf route full of blanks. The ninth race is a turf route, an allowance — not a maiden — with ten starters as filed. It hits Cell A on its class and Cell B on its field size before you read a single past performance: under the lab's rule it files no pick, full stop, and the reasoning you do next is for your own education, not to overturn the cells. Two of the ten are Unknowns, one of them a shipper the morning line makes third choice. The favorite's pace shape depends on pressure you cannot verify, because one of the possible pace sources is a blank. Your sheet knowledge of this race is: a surface-and-shape cell with a counted losing return in its two overlapping forms, a model that declares itself not covering the field, and a price structure resting partly on a horse you cannot score. The pass writes itself. What it costs: the chance that your eye for turf routes beats a large selection count, exercised on a field you partly cannot read. What it saves is priced in the arithmetic below. You watch the race. Watching is free, and the race still teaches you the shipper for next time — the one useful thing a passed race pays, it pays in information.

The favorite from the other surface. A dirt sprint makes a horse the clear favorite off three strong figures, every one of them earned on turf. No filing rule in this chapter excludes the race; it is a dirt race, the field is readable, nobody is Unknown. The V4 screen flags the favorite and tells you its measured

context: favorites in this shape won less often and returned less than favorites generally, and the screen separated the same way on the control population. Your choices are honest ones. You may bet another horse — but only if that horse has its own case, because the screen gives you no candidate, and "not the favorite" is a field of seven horses, not a pick. You may bet the favorite anyway, if you hold a reason the screen cannot see: a workout pattern you trust, a trainer whose surface switches you have counted yourself, a price generous enough to pay for a one-in-four chance with room to spare. Or you may pass, on the stated ground that the race's central price is built on evidence from different ground and the rest of the field has not made its own case to you. Notice what the pass is not: it is not a bet against the favorite. It does not need the favorite to lose. It needs only your judgment that no price in this race pays you for the risk you can actually assess — and on this race, the risk you can assess worst belongs to the horse everyone else is betting most.

The reach in the last race. Your day has gone badly, which is when this race arrives: a readable dirt field, a horse you half-like at a morning line above 8-to-1, one figure that fits, a story that assembles itself the way late-day stories do. The lab's ceiling answers before the story finishes: no leg files above 8-to-1, not because the horse cannot win, but because a disciplined one-pick operation counted where its selections earn their keep and drew the line where the reaching starts. Your ceiling, set at breakfast, does the same work. The pass here saves nothing measurable on this one race — any single race is one draw from a distribution, and this horse may win at a handsome price and be remembered at your table for a year. What it saves is the pattern: the late-day reach is a recurring expense in every casual

player's season, and a rule that fires before the story starts is the only version of willpower that survives the eighth race. Add the fourth shape yourself, because it needs no figure and this chapter attaches none to it: the race where your signals point four ways at once — pace says one horse, figures another, the market a third, and the bias line from Chapter 16 declines to lean. Disagreement among your own tools is not a puzzle to solve by betting anyway. It is the tools reporting that the race, for you, is unread at the standard your money deserves. Pass it, and let the absence of a lab count for that shape bother you exactly as much as it should: it is general reasoning, stated as such, and it will stay that way until somebody counts it.

What passing costs and saves across a season

Now count the habit. This chapter makes a choice for its season arithmetic, and it states the choice here: it uses the rebuilt favorite figure, with its population label, instead of restating the whole section with no numbers at all. That keeps the passage honest about size — the rebuilt record is three days, not a season — and it keeps the arithmetic tied to a publishable count.

The rebuilt figure is this. On the lab's own settled product record: the filed cards of November 2–4, 2026 — 111 races, the morning-line favorite started in 101 races, had 27 winners, a win rate of 26.7%, and returned $111.10 on $202.00 staked, for an ROI of –45.00%. Bet the favorite blindly at that rate and one hundred races at $2 is $200 staked and, at that rate, $90 lost. That rate comes from three days. Treat it as a small, swingy sample, not as a long-run favorite rate.

No rebuilt figures exist in this rebuild for the §25 turf cells or for the V4-flagged favorites. This chapter therefore restates those parts of the arithmetic qualitatively. In their original counts, the named turf cells and the flagged favorites lost more than the general baseline they were compared with. Passing a race in one of those shapes moves its expected line from a negative figure to zero. The chapter does not invent a rebuilt dollar saving for shapes it did not recount.

And passing one hundred races in any shape costs, in stakes, exactly zero dollars, and returns exactly zero dollars, for a season line of $0.00 where a betting line at the rebuilt favorite rate reads –$90.00 on $200 staked.

State the other side with the same plainness, because a book that only counts savings is selling something. Passing forgoes the winners. At the rebuilt favorite win rate, about 27 of those 100 passed favorite races are won by the horse you declined, and some of them pay well, and you will remember several by name. Passing also forgoes any race in which your personal reading genuinely beats the population rate — such races exist for every serious player, and a pass rule drawn too wide will refuse some of them, which is why this chapter's rules are drawn at cells and shapes with counts behind them, and why the wide version of every rule in it was the version the tests refused. And passing is not a profit. Zero is the return on a passed race, every time, without exception. The claim for passing is narrower and, counted across a season, large enough: the baseline activity of betting loses money, several named shapes lose more than that in their counts, and every race you move from the betting column to the watching column moves its expected line from a

negative figure to zero. A player who passes races that fall in the named shapes has not won anything. He has kept his stake in his pocket to bet on races he can read, at prices he can defend.

That is the standard to hold every future pass rule to, including your own. A pass earns its place the way a pick does: a shape defined in advance, a count over every race in the shape, a comparison against the baseline it replaces, and the honesty to say which kind of claim supports it — a counted return, as with the turf cells; a coverage refusal, as with the Unknown; a filing discipline, as with the ceiling; or an association, labeled as one, as with the screen. Four different grounds, four different strengths, one habit. The races to leave alone are not the races you know nothing about. They are the races you know exactly one thing about — and that one thing is that they are not yours to bet.

The three lessons of the pass

First: a blank is not an average. A horse with no usable data has no probability you are entitled to use — not zero, not the field mean. A race your tools declare uncovered is a race you may decline on that ground alone, and the declaration is the honest output, not a failure of the tools.

Second: pass cells, not surfaces, and rules, not hunches. The counted evidence refused a ban on turf and supported two cells inside it; it refused a ban on the unknown horse's race and supported refusing the model's coverage of it. The narrow, verified version of a pass is worth more than the sweeping one, every time it has been tested.

Third: passing is a position with a return of zero, and zero beats the measured alternatives. Against baselines that lose a fifth of the stake and named shapes that lose a quarter, the pass holds its value — not because it wins, but because its return is known before the gates open.

Chapter 18 — Money: units, stops, and records

Every chapter before this one has been about the horse side of the game: which ideas describe a race, which ones pick winners, which ones lose less than the favorite and which ones lose more. This chapter is about the money side, and it starts with a confession that protects everything else in it.

The rules in this chapter are not findings.

Nothing in the seventeen chapters ahead of this one tested a staking plan. No lab experiment compared one unit size against another on thousands of races and came back with a verdict. What follows is the operating practice of one laboratory that bets on paper, written down as a guide for a $1,000 bankroll: conservative guidelines, chosen in advance, designed to do a limited job. Read the job description carefully, because a staking plan that promises more than this is lying to you:

Staking and stopping rules manage the size and speed of losses. They do not change a bet's expectation. No staking plan turns a negative-expectation method into a positive one.

That sentence is the whole chapter in advance. The rest is arithmetic and habit.

What a staking plan can and cannot do

A bet's expectation is set before money management enters the room. It is set by the horse, the price, and the takeout. If a method returns about 62 cents for every dollar it bets over a run

— and the rebuilt headline rate in this chapter, –38.24% on the lab's own settled product record: the filed cards of November 2–4, 2026 — 111 races, sits in that neighborhood — then every dollar you push through that method costs you about 38 cents on average, whether you push it through in $2 pieces, $20 pieces, or a pattern you read about in a forum at 2 a.m. Three days is a small, swingy sample; the chapter uses the rate to price the arithmetic, not to claim a long-run rate.

Changing the size of the pieces changes three real things:

- **How fast the expected loss arrives.** Bigger pieces, faster loss. Smaller pieces, slower.
- **How lumpy the ride feels.** Variance — the swings around the average — scales with stake size. The same bad day is a shrug at one unit size and a crisis at another.
- **Whether you survive your own worst stretch.** Every method, good or bad, has losing runs that are longer than a new player expects. A staking plan decides in advance whether a normal bad stretch ends your season or just dents it.

That is the full list. A staking plan does not change which horses win. It does not change what they pay. It does not compound a losing method into a winning one, because compounding a negative rate produces a larger negative number, faster. The systems that claim otherwise — double up after a loss, press after a win, bet a percentage that "locks in" profits — all rearrange the same losses into different shapes. Some shapes feel better for a while. One common shape, the chase, concentrates the whole season's loss into a single afternoon. The arithmetic underneath never moved.

So why bother with rules at all? Because the casual player's real enemy is rarely the average. It is the sequence. Losses do not arrive evenly spaced, like fence posts. They arrive in bunches, on the days you played the biggest, with your judgment at its worst. A staking plan written in advance is a set of decisions made by the calm version of you, binding on the tilted version of you. That is worth having. It is also all it is.

Hold that frame and the rules below read correctly: not as a way to win, but as a way to know in advance what following your method costs, to keep any single day from doing outsized damage, and to stop on rules instead of on emotion.

The unit: bet a fixed, small piece

The guide's first rule:

The unit is 0.2% of the bankroll, recalculated once a week, with a floor of $2.00.

On a $1,000 bankroll, 0.2% is $2.00 per bet. Two dollars. If that number feels anticlimactic, good — that reaction is the reason the rule exists. A unit is not sized to make a winning day thrilling. It is sized to make a losing month survivable.

Here is the same typical day at three unit sizes, from the guide's own table. A full day of play in the lab's book runs about 35 bets. One real bad day in the lab's record — the 2026-11-02 card, which returned $44.18 on $78 at the $2 base — lost $33.82 at base stakes. Scale that day up:

Unit	Share of a $1,000 roll	35-bet day staked	The same bad day costs	Loss as share of roll
$2	0.2%	$70	−$33.82	−3.4%

Unit	Share of a $1,000 roll	35-bet day staked	The same bad day costs	Loss as share of roll
$5	0.5%	$175	−$84.55	−8.5%
$10	1.0%	$350	−$169.10	−16.9%

One ordinary bad day — not a disaster, just a bad day of a kind that happens several times a season — takes a sixth of the bankroll at a $10 unit. At $2 it takes a thirtieth. Nothing about the horses changed between those rows. Only the damage did.

The guide sets $5 as the absolute ceiling on a $1,000 roll, and only for a player following every other rule in this chapter. The standard play is $2. If your bankroll is smaller, the percentage matters more than the floor: the guide's $2 floor exists because its own books are kept at a $2 base, but on a $400 roll a flat $2 unit is 0.5% a bet, and a 35-bet day stakes $70 — 17.5% of everything you have. Run the percentage on your own roll before you copy the dollar figure. The rule is the 0.2%. The $2 is what 0.2% happens to equal at $1,000.

Two habits ride along with the fixed unit, and both exist to kill a specific temptation:

Every bet stakes the same unit. There is no bigger bet on the ones you like. The moment some bets are $2 and others are $15 "because this one is different," you no longer have a staking plan — you have your afternoon mood, denominated in dollars. Your records (the last section of this chapter) also become unreadable, because you can no longer tell whether the method worked or the sizing worked. One unit, every bet, is what makes the record a record of the method.

Winnings are not house money. The phrase "playing with the house's money" has cost casual players more than any bad beat in racing. A dollar you won on Saturday is your dollar on Sunday, worth exactly as much as the dollar you brought on Friday. The guide's reset rule below treats it that way automatically. Your language should too.

The daily cap: the day has a budget, not just the bet

The guide's second rule:

Never stake more than 10% of the current bankroll in one day — $100 at the start on a $1,000 roll.

Notice what this rule does that the unit rule does not. The unit rule sizes each bet. The cap sizes the day. A player can follow a sensible unit and still have an unbounded day: 35 bets becomes 60 because the card was interesting, then a nightcap track opens, then a carryover pool appears. The cap ends the day arithmetically. At the $2 unit, a normal day stakes $70 to $78 and sits inside the cap without trying. The cap binds only on unusually large days — in the guide's terms, cards of 51 or more bets.

When the cap binds, the guide's instruction is specific, and it is the opposite of what most players do: **scale the whole day's unit down to fit. Never drop selected races to fit, and never scale up.** Dropping races to fit turns the day into a different, untested bet — you kept the races you liked and cut the ones you didn't, which is opinion wearing a budget's clothes. Scaling the unit down keeps the method intact and shrinks only the money.

Where a track's minimum stake prevents scaling below the base, the day plays at base and the exception is written down for what it is.

The cap has a second, quieter function. It makes the worst legal day knowable in advance. Before the first race, a player under these rules can say: "The most today can stake is $100, and a total wipeout — every bet losing, which will essentially never happen — costs me a tenth of my roll." Most days lose a fraction of what they stake. A player who knows the ceiling can watch a bad afternoon without the background dread that the losses might be unbounded. Dread is what makes people chase. The cap removes the reason for it.

What a season costs: read this before the first bet

A staking plan for a method with a negative expectation is a budget, and a budget should be stated as a number before the season starts, not discovered at the end of it. The guide does this with the rebuilt headline rate. The headline book (official legs plus counter-move legs, as published nightly) on the lab's own settled product record: the filed cards of November 2–4, 2026 — 111 races, returned −38.24%: 108 legs, 28 winners, $216.00 staked, $133.40 returned. Keep the warning with the rate wherever it prints: three days is a small, swingy sample. It is not a long-run measurement. At that rate, every $2 leg returns about $1.24 and costs about $0.76. Staked at the $2 unit, about 35 legs a day, the expected cost runs:

Horizon	Staked	Expected loss at −38.24%
One day	~$70	~$26.60
One week (7 days)	~$490	~$186

Horizon	Staked	Expected loss at −38.24%
One month (30 days)	~$2,100	~$798
Expected runway of the $1,000	—	about 5–6 weeks (roughly 37 days at $26.60 a day)

Read that table the way the guide insists it be read. It is an average over a run priced at a three-day rate. Real months arrive lumpier than averages. Variance decides the order of the days; the expectation decides where the roll is heading. At the rebuilt rate, daily full-card play on $1,000 spends the roll in about five to six weeks. That is not a warning label pasted onto the plan. That *is* the plan, stated honestly: about $26.60 a day of expected cost for a day of full-card play at that rate, the way a round of golf or a concert ticket has a price.

Your own method's rate will differ from this rate, and Chapter 19 shows you how to measure it. Run this same table with your number before you set your unit. If the monthly figure is more than the entertainment is worth to you, the guide's answer is blunt and correct: the right stake is zero, or the right frequency is fewer days. **Play fewer days, not larger ones.** A player who plays Saturdays only at the $2 unit cuts the monthly cost to a fraction without touching the method. A player who "makes Saturdays count" with a $10 unit keeps the cost and quintuples the damage of every bad one.

Resizing: once a week, in one direction when it matters

Bankrolls move, so units must move — but on a schedule, not on a feeling. The guide's reset rules:

Units reset once a week — Monday, before the day's play — to 0.2% of the actual bankroll that morning, rounded down to the nearest $0.50, with the $2.00 floor.

No mid-week changes, in either direction. A winning Tuesday does not raise Wednesday's unit. A losing Tuesday does not lower it. Mid-week resizing is how a staking plan becomes a mood ring: raise it while confident, which is usually near a local peak, and cut it while discouraged, which is usually near a local bottom. The weekly reset is deliberately boring. Boring is the point.

Then the rule that gives the plan its spine:

After the bankroll has touched −20% — $800 on a $1,000 start — units may only ratchet down. The weekly reset still runs. It still recalculates 0.2% of the balance. But it may never set a unit higher than the previous week's, for the rest of the season. A bankroll that has fallen a fifth has delivered information: the season so far costs at least what the method's rate predicted, possibly more. The ratchet accepts that information once, mechanically, without a meeting about it.

Worked through a real day, the system looks like this. The guide replays the lab's 2026-11-02 card under its own rules: bankroll at the reset, $1,000, so the unit is $2.00. The card carries 39 bets — 37 races, with a handful of races carrying a second, separate bet by the method's design — for a stake of $78, inside the $100 daily cap. The card settles with 12 winners and $44.18 returned: a loss of $33.82, or 3.4% of the roll. No tripwire fires; the day is bad and unremarkable. The bankroll is $966.18. The unit stays $2.00 until Monday, when 0.2% of the new balance — about $1.93 — rounds against the $2.00 floor, so the unit remains

$2.00. Notice how little there is to decide in that paragraph. Every number was fixed before the day started. A bad day, absorbed exactly as designed: noticeable, not damaging, and nothing to make a decision about while annoyed.

The drawdown ladder: decide the stops while you are calm

Every player has a stopping rule. The only question is whether it was written in advance or invented at the worst moment. The invented ones sound like this: "I'll quit when it stops being fun," which translates, in practice, to quitting after the loss that finally hurts — the maximum damage point, chosen by pain. The guide replaces pain with a ladder, decided from the starting $1,000 before the season's first bet:

At −20% ($800): review. Nothing stops. One thing changes — the ratchet from the previous section engages, and units may only move down from here. The guide is explicit that this is a checkpoint, not an alarm: a method returning at a clearly negative rate, such as the rebuilt headline rate of −38.24% on the lab's own settled product record: the filed cards of November 2–4, 2026 — 111 races (three days, a small, swingy sample), is *expected* to visit −20% on a long enough run. Arriving there means the season is going roughly as priced. The review is a moment to check the records (next section), confirm the bets being made still match the method being tested, and continue at the same or smaller size.

At −35% ($650): halve the exposure. Do not change which horses the method selects. Change how much of the method you play. The guide offers two equivalent forms: drop the

method's secondary bets and play its main bet only, or play alternate days. Either one halves the daily stake while leaving the selection method untouched, so the record stays interpretable — you are still measuring the same idea, at half throttle. What you may not do at –35% is the thing every instinct suggests: change methods. A new method adopted at the bottom of a drawdown starts its own record from zero evidence, and the drawdown that prompted the switch gets quietly buried. The ladder exists so that the response to losses is smaller exposure, not a new opinion.

At –50% ($500): stop. The season is over. The guide's reasoning is worth stating in its own terms: reaching –50% means the method has run at or below its measured rate for a long stretch. Continuing past that point is no longer "following the selections." It is funding a measured negative expectation with no new information. The remaining $500 is not a betting bankroll anymore. It goes back to being money.

The bad-day tripwire: any single day that loses 5% or more of the bankroll puts the next day at the floor unit, and that week's reset may not raise the unit. In the unit table above, the 2026-11-02-shaped bad day trips this wire at a $5 unit (–8.5%) and at a $10 unit (–16.9%), but not at the $2 unit (–3.4%). Read that as the ladder grading the unit sizes for you: at the guide's unit, a bad day is weather; at three times the unit, the same day is an event that changes the rules for a week.

Two features of the ladder deserve emphasis, because players get both wrong.

First, the percentages are of the **starting** bankroll in the guide's presentation — $800, $650, $500 are fixed landmarks on a $1,000 start, not moving targets recalculated off a shrinking balance. Fixed landmarks cannot be argued with mid-season. A moving "20% below wherever I am" line can be rationalized downward forever; $500 cannot.

Second, none of the rungs promise that stopping saves a method that would have recovered. Some seasons that touch –35% would have clawed back. The ladder does not know, and does not claim to know. What it knows is the price of finding out: at –50%, finding out costs the second half of the roll. A stopping rule is insurance, and like the place and show bets discussed elsewhere in the lab's work, insurance has a premium — sometimes you pay it on a season that did not need it. The premium is the cost of never having the season that takes everything. Players who refuse all premiums eventually collect the full loss instead.

Scale the ladder to your own start with the same percentages: on a $500 roll, review at $400, halve at $325, stop at $250; on $2,000, review at $1,600, halve at $1,300, stop at $1,000. Write your three numbers on the first page of your record book before the first bet. If you cannot watch those numbers approach without flinching, the bankroll — or the frequency of play — is too big, and the time to learn that is now, at full strength, not at the second rung.

Records: the section that makes the rest of the book possible

Everything in this book exists because somebody wrote things down. The favorite's record in Chapter 1, the verdicts in Chapters 11 through 15, the worked example in the next chapter — all of it is records, counted later. The lab's habit is not a talent. It is a notebook, kept without exception, and it is the one practice in this chapter that is not a guideline. It is a requirement. **Keep records, or stop pretending to learn.**

The reason is simple: memory is a publicist, not an accountant. Every player remembers the $38 winner in the ninth race. Almost nobody remembers, accurately, the eleven bets before it. Ask a player in December how his year went and you will get a story. Ask his notebook and you will get a number. The story is always better than the number, and the gap between them is exactly the size of the lesson he did not learn.

Here is what to write down after every bet. It takes under a minute per bet, and it is the same list whether you use a spreadsheet, a notebook, or the back of the program you will inevitably lose — so use the spreadsheet or the notebook.

The bet itself: - Date, track, and race number. Without these, the bet cannot be checked against the official chart later, and unchecked records drift. - The horse's name and program number. Names get misremembered; program numbers do not. - Bet type and stake: win, place, show, exacta — and the exact dollars. "About $10" is not a record. - The price: the horse's morning-line odds and, when you can get them, its final odds.

The gap between the two is data about your timing and about the crowd, and you cannot study it if you never wrote either number down.

The outcome: - Finish position, and the official payoff if the bet cashed. Write $0 for a loser, in figures, not a dash. Dashes get skipped when you total the column. Zeros do not. - Amount returned, and the running bankroll. Update the running total the same day. A bankroll figure that is three weeks stale is how a player reaches a stop rung without noticing he passed it.

The reasoning — the part everyone skips, and the part that matters most: - The rule or method that produced the bet, named specifically: not "my system," but the rule as you would state it to a stranger. If you follow a published selection method, write its name and version. If the bet came from your own idea, write the idea in one line. - The edge you believed you had, in one sentence, written *before* the race if at all possible and immediately after at the latest. "Best late figure in a field full of early speed" is a reason. "Liked him" is not — and writing "liked him" a dozen times in a season is itself a finding, if you are honest enough to write it. - Anything unusual you knew at bet time: a scratch that changed the race's shape, a surface change, a horse you meant to bet and didn't. These notes are what let you audit yourself later for the two great record-keeping sins: quietly dropping the bets that embarrassed you, and quietly adopting the winners you only thought about.

Then, on a schedule — the guide's weekly reset is a natural moment — total the record and read it the way the lab reads its own: number of bets, total staked, total returned, return on investment, and the same figures for the favorite in the same races if you can reconstruct them. Compare like with like, over

the whole count, not the memorable subset. Your notebook will tell you things your memory will not: that your win bets outperform your exotics by twenty points, that your Sunday plays lose at twice your Saturday rate, that the last three "one-line reasons" that said *drop in class* all lost. Some of those patterns will be noise — Chapter 4 explains how easily noise dresses as pattern, and small counts prove little. But you cannot even ask the question without the record. With it, every season ends with evidence instead of a story.

One rule governs the notebook itself, and it has no exceptions: **write down every bet, including the ones you wish you hadn't made, at the size you actually made them.** A record with the bad Tuesday removed is not a record; it is advertising. The lab's results in this book count every race a rule touched, including the humiliating ones, for the same reason. Your notebook is your lab. Run it the same way.

The rules on one page

For the casual player who plays a method — this book's, the lab's, or their own — at a $1,000 bankroll:

1. **Unit:** 0.2% of the bankroll — $2 per bet at $1,000 — the same on every bet, no favorites-within-the-favorites.
2. **Daily cap:** never stake more than 10% of the current bankroll in a day. If the day does not fit, shrink the unit, never cherry-pick the races.
3. **Reset:** once a week, from the actual balance, rounded down. No mid-week changes for any reason. Winnings are bankroll, not house money.

4. **Ratchet:** after the roll touches –20% ($800), units may only move down.
5. **–35% ($650):** halve the exposure — fewer bet types or fewer days — without changing the method.
6. **–50% ($500):** stop. The season is over and the remainder is not a bankroll.
7. **Bad day (–5% of the roll in one day):** next day at the floor unit; that week's reset cannot raise the unit.
8. **Records:** every bet, every field above, totaled weekly. No exceptions, no deletions.

And above the list, the sentence the list cannot replace: these are conservative operating guidelines from the lab's practice, not tested findings. They will make a losing method lose more slowly and a bad day hurt less. They will not make a losing method win. If you want the money side of this game to work in your favor, the only lever that has ever moved expectation is the one the rest of this book is about — the quality of the bets themselves, tested the way the final chapter shows you how to test them.

Chapter 19 — Test it yourself

This book ends where it began: with a count.

You have read eighteen chapters of other people's tests — mostly ours. Some ideas survived their counts and most did not, and if the book has a single lesson it is that the counting, not the idea, is what separates a handicapper from a storyteller. So the last chapter hands you the count. What follows is the lab's standard, reduced to a kitchen-table method you can run at your own track, on free charts, with a pencil and a spreadsheet. Then it walks through one complete lab test exactly as it was run — rule, denominator, comparison, verdict — so you can see the method filled in before you are handed the blank version.

The method has five parts. None of them is optional, and the order matters.

Part 1: Write the rule in advance, in full

Before you look at a single result, write down the idea as a rule a stranger could apply without asking you anything. The lab freezes its rules in a written preregistration before any outcome is read, and the freeze covers five things:

- **The rule itself.** Exactly which horse gets picked, in language with no judgment calls left in it. "Bet closers in pace meltdowns" is not a rule. "In fields of nine or more, bet the horse with the highest late-pace figure among horses whose early-pace figure is not the field's best" is a rule — a stranger holding the same past performances picks the same horse you do.

- **The population.** Which races the rule applies to: which tracks, which dates, which race types. "All races at my track from January through June" is a population. "Races where the rule looks right" is not.
- **The exclusions.** What gets skipped, stated before the fact: races missing a figure the rule needs, races where the pick is scratched, races taken off the turf — whatever applies. Exclusions invented after the results are known are how losing tests get edited into winning ones.
- **The benchmark.** What the rule must beat. In this book the benchmark is almost always the simplest one in racing: a $2 win bet on the favorite in the same races. Your rule does not have to beat the takeout to be interesting — Chapter 1 settled that — but it has to beat doing the obvious thing, or the obvious thing is better than your rule.
- **The margin.** How much better than the benchmark counts as "worked," decided in advance. A rule that returns one tenth of a point more than the favorite over 200 races has not worked; it has tied, within noise. The lab's gates, shown in the worked example below, are one way to set a margin. For a kitchen-table test, a plain version serves: name the return gap you would need to see, and the minimum number of races, before you start.

Write all five down. Date the page. Do not edit it afterward; if the rule needs changing, that is a new test with a new page, and the old page stays in the file with its result. This feels bureaucratic right up until the first time a rule you loved fails and your memory starts negotiating. The dated page does not negotiate.

Part 2: Count every race it touches

Apply the frozen rule to the whole population. Every race. The races where the rule's horse went off at a silly price, the races where it rained, the races where you are now sure the rule "wasn't meant" to apply — if the frozen page does not exclude them, they count.

This is the part of the method that does the real work, and the part most private tests skip. A player tests an idea by remembering it when it wins. The count remembers it when it loses, too, and there is no other difference between a test and an anecdote. The lab's phrase for the denominator is worth borrowing: every verdict in this book states its race count next to its result, because a return figure without a count is a rumor. A losing return over a few dozen races is weather. The same return over many thousands of races is climate.

Keep a result file as you go — one row per race: date, track, race number, your rule's horse, the favorite, the winner, what your horse paid if it won, what the favorite paid if it won. Chapter 18's record list covers the same ground from the bettor's side. Build the file while the races are fresh; reconstructing it from memory a month later rebuilds the bias you are trying to remove.

Part 3: Compare against just betting the favorite

When the population is counted, settle three columns at $2 win stakes: your rule, and the favorite in the same races. Report each column's races, win rate, money staked, money returned, and return on investment.

Compare on **identical races**. If your rule skipped a race because its figure was missing, the favorite's column skips it too. A comparison where each side plays a different set of races measures the sets, not the methods. The lab calls this the paired comparison, and it also computes a stricter version of it: the average per-race difference between the two columns, with a range around it (a confidence interval) that says how much of the gap could be luck of the ordering. You can approximate the same honesty without the statistics: if your rule's edge over the favorite is small enough that swapping a handful of winners would erase it, say so in the file, in those words.

Also count the **duel**: in the races where your rule and the favorite named different horses, how often did yours win, and how often did the favorite win anyway? A rule can post a better return than the favorite while losing the duel badly — that means its winners paid more, not that it picked better — and the two claims are different, as Chapter 5 established. Write both down. Grade both.

Part 4: Keep the result file, whatever it says

The file is the deliverable. Not the profit — there usually is not one — but the file: the frozen rule page, the per-race rows, the totals, the comparison, and the verdict you reached, dated. A test you ran and filed teaches you something every time you reread it. A test you ran in your head teaches you whatever you currently wish were true.

Part 5: Let it die if it loses

The standard ends with the step that makes the other four worth doing. If the rule fails its margin against the favorite, it is not "promising." It is not "one tweak away." It is dead, at the population and margin you named, and the file records the death. You may test a genuinely different rule next — a new page, a new freeze — but the dead rule does not get quietly revived with its worst races reclassified. The lab's closed list, printed as Appendix A of this book, is this step practiced at scale: dozens of ideas, each with its verdict, none of them still on the payroll.

Now watch the whole machine run once, on a real test, with real numbers.

The worked example: does the most efficient closer beat the favorite?

The idea. Chapter 11 covers pace efficiency — the notion that a horse which keeps most of its early speed at the finish (a high late figure relative to its early figure) is doing something physically real that the crowd underprices. As a pick rule, it says: forget the favorite; bet the most efficient horse in the race.

The freeze. The lab wrote this test down on November 5, 2026, in a preregistration frozen before any outcome was read, as part of a six-method program. Here is the freeze, translated line by line into the five parts above:

- **Rule:** For each starter, compute its efficiency ratio — its late pace figure divided by its early pace figure. Bet the horse with the highest ratio. Ties go to the horse the

morning line ranks higher. (The lab's sheet names this rule R-PEI, and the ratio PEI in its corpus form. The test bed carries no turn figure, so the ratio uses the early figure as its base; a second form using a turn figure was scheduled and tested separately, on a smaller set of the lab's own cards, with the same verdict.)

- **Population:** The lab's primary test bed — a full population of races from one year at the tracks its product covers. A second, larger population from the same year across all tracks in the records was named in advance as the control, to be reported alongside.
- **Exclusions:** Any race in which no starter carries a morning line cannot be scored, because the benchmark cannot be named in it. The freeze's conventions also fixed the small print before the fact: bets are $2 win bets; a winning pick returns its official win payoff and anything else returns $0; a race whose winner's payoff is missing counts the payoff as $0 rather than being dropped.
- **Benchmark:** The morning-line favorite in the same races — the ML-BASE column this book has used since Chapter 1.
- **Margin** (**the gates**)**:** The rule counts as SUPPORTED only if all three hold on the primary population: (1) its paired per-race edge over the favorite is positive with the bottom of its 95% confidence interval above zero — the edge must be too big to be plausibly luck; (2) it wins the duel — in races where it and the favorite differ, its horses win more often than the favorite does; (3) its return on investment beats the favorite's return. Anything less is recorded as NOT SUPPORTED. Those are demanding gates,

deliberately. A rule that cannot clear them has not earned a bettor's money, whatever its story.

The count. The evaluation ran on nearly the whole population. The few missing races are named in the report, and the reason is the frozen exclusion working exactly as written: a race survives in the records without any starter-level morning-line values, so no favorite exists to compare against and no rule can be scored. Missing races, named, with their reasons — that is what an exclusion looks like when it is honest. In one further race no starter carried both pace figures, so the rule fell back to the favorite there, as its conventions provided.

The result. Read the win rate first. The most efficient horse in the race won far less often than the favorite did — about half as often, in the lab's summary. That single line ends most versions of the idea: efficiency, as a way of naming winners, names them far worse than the crowd does. The return column was closer — the rule's winners paid better prices, which clawed back most of the gap — and the paired comparison still placed the rule behind the favorite per race, with a confidence interval that included zero and included positive territory. That interval alone means the return gap is too small, relative to its noise, to convict or acquit on its own. The duel settles the picking question: across the races where the rule and the favorite named different horses, the favorite won far more of those head-to-head races than the rule did, by a wide margin. There is no reading of that result that flatters the rule.

The control. The same frozen rule, run on the larger control population, returned the same answer: the rule won far less often than the favorite, returned less, ran behind the favorite in

the paired comparison, and lost the duel widely. A result that flips when the population widens was probably measuring the population. This one did not flip.

The verdict. Against the three frozen gates: the interval's bottom is below zero — fail. The duel is negative — fail. The return sits below the favorite's — fail. **R-PEI: NOT SUPPORTED.** The rule was not adopted, not combined with anything, and not retuned. It went on the closed list, and the diagnostic scraps worth keeping — measurements of how efficiency relates to favorites, covered in Chapter 11 — were filed as description, not as a pick rule. That distinction is the lab's whole discipline in one gesture: a test can teach you something true about races and still forbid you to bet its rule.

Notice what the example did not require. No private data: the inputs are pace figures of the kind printed in ordinary past performances, and the test bed is a corpus of past results. No live betting: every leg is a paper $2. No large bankroll, no software beyond a spreadsheet's worth of arithmetic, and no discretion anywhere in the middle — the freeze removed discretion on purpose, because discretion is where failed rules go to be rescued.

Three ways this test could have been cheated — and what the freeze did about each

It is worth pausing on the worked example to see the failure modes it was built to survive, because your own tests will meet all three.

The rescue edit. Halfway through the count, a tester notices the rule does terribly in turf races, or in sprints, and quietly narrows it. The narrowed rule posts a better number, and the tester reports the better number as the test. Nothing about that number is evidence: the population was chosen *because* the rule won there. The freeze blocks the rescue by dating the population page before the count. If turf races look different afterward, that observation is a new idea, and it gets a new freeze and a fresh count on races the first test never touched — or, at minimum, an honest label as a pattern found inside old data, unproven until fresh races judge it.

The shrinking denominator. A tester drops races one at a time, each with a reason that sounds reasonable in isolation: the pick was scratched late, the track came up sloppy, the figures looked unreliable that day. Each drop is small. Together they remove a slice of the losses, and the slice is never random — races get scrutinized for defects when they lose. The worked example's denominator discipline is the antidote: nearly the whole population counted, with each missing race named and its reason printed. When you finish your own count, write that same sentence — "I counted N of M races in my population; the missing ones are listed here, with reasons" — and make yourself list them. A denominator you have to defend in writing stays honest.

The small count that shouts. The same report carries a warning in the other direction. One of its vulnerability screens — a rule meant to flag favorites whose lone early speed came with a weak late figure — flagged only a few dozen favorites, and those favorites won at a high rate. Read naively, the screen "discovered" a strength. Read at its count, it discovered nothing:

a screen that flags only a few dozen favorites cannot carry a strong claim either way, and the lab's report says exactly that about it. Your tests will produce numbers like this — a rule that goes well in its first month at your track. A short winning run is a story. The freeze's minimum race count exists so the story has to finish before it gets a verdict. Set yours high enough to hurt a little. A test you can finish in a weekend is a test whose answer is mostly luck.

None of these cheats requires dishonesty. Each one is a reasonable-sounding decision made at the moment it does the most damage, by a person who wants the idea to be true. That is why the standard is mechanical. You do not rise to the level of your intentions when a favorite idea is losing. You fall to the level of your paperwork.

Your turn: the blank worksheet

Here is the same test with the numbers removed. Copy these pages — literally, into a notebook or a document — and fill them in for one idea of your own, at one track you actually follow, before you count a single race.

Step 1 — Freeze the rule (before any results). - My rule, in one paragraph a stranger could apply: _____ - The figures or facts it uses, and where I get them: _____ - Tie-breaks and fallbacks, for when the rule's first choice is missing or tied: _____ - Population: track(s) _____, dates _____ to _____, race types included: _____ - Exclusions, stated now: _____ - Benchmark: the favorite in the same races, $2 win bets. (If you choose a different benchmark, write it and say why:

_____) - Margin: my rule counts as supported only if, over at least _____ races, it beats the favorite's return by at least _____ points **and** wins the head-to-head duel. Confidence range: I will treat any edge small enough that swapping _____ winners would erase it as unproven, and I will write that sentence in the result file either way. - Date frozen: _____. Signature, if you like — the point is the commitment, not the ceremony.

Step 2 — Get the races, free. You need, for each race in your population: the field with the figures your rule uses, the morning line or final odds, the finish order, and the win payoffs. Official result charts are published free by the major chart services and by most tracks, and they carry finish order and payoffs; past-performance lines for horses in upcoming races appear in the free past performances tracks and handicapping sites post for each card, and a horse's lines in those sheets include its earlier races at your track. Work backward from the charts for results and forward from the sheets for figures, and keep every race in the population — the slow, methodical completeness of the count is the test. If a race's data cannot be found, write it in the file as missing, with the reason, the way the lab's report names its unscorable races. Missing races are a fact about your data. Silently skipped races are a fact about your honesty.

Step 3 — Count. One row per race: date, track, race number, my rule's horse, the favorite, the winner, my horse's payoff (or $0), the favorite's payoff (or $0). Apply the rule exactly as frozen, especially in the races where you now believe it is wrong. Finish the whole population before you total anything. Peeking at a running total and adjusting the rule mid-count converts the test

into a fitting exercise — you will end up with a rule perfectly shaped to races it has already seen, which Chapter 4 explains is no evidence at all about the next ones.

Step 4 — Settle and compare. Total both columns on the identical races: bets, winners, win rate, staked, returned, return. Compute the per-race difference between the columns and look at its size against the count — a two-point edge over 60 races is a coin flip wearing a lab coat; the same edge over 2,000 races deserves a second look. Count the duel. Then grade the rule against the margin you froze in Step 1 — not against a kinder margin invented now.

Step 5 — File the verdict. Write the result page: the rule as frozen, the population as counted, the missing races and why, both columns' full numbers, the duel, and one of two verdicts — SUPPORTED at my stated margin, or NOT SUPPORTED — with the date. File it where you will find it next season. If the verdict is SUPPORTED, the correct next move is still not a larger bet; it is a second count, on new races the rule has never seen, under the same freeze. A rule that survives fresh races has earned a prospective record. A rule that only survives the races it was born in has earned a file page.

Step 6 — If it lost, close it. Write the idea on your own closed list with its record, the way Appendix A lists the lab's. Do not delete the file. The closed list is not a wall of shame; it is the asset. Every idea on it is a tuition payment already made, and its value is that you never pay it twice. The player across the aisle from you will test the same idea next spring, from a tip, at full stakes, with no freeze and no count. You will know its record because you kept it.

The standard, one last time

This book opened with the favorite: the best handicapper you know, losing a fifth of every dollar to the takeout and the crowd's own sharpness. It closes with the only honest response the lab has found to that fact. Not a secret — the book has published its best method's full record, and the record loses. Not surrender — the counts did find real structure in races, measurements that describe favorites honestly, screens that separate, a consensus that loses a few cents less than the chalk. The response is a standard, and it fits in one paragraph:

Write the rule before the races. Count every race it touches. Compare it to the favorite, on the same races, at the same stakes. Keep the file. And let it die if it loses.

Everything else in handicapping — the figures, the shapes, the class pars, the bias reads, the money rules that keep a bad month from ending you — is material to feed that standard. An idea that has passed it is worth your $2. An idea that has not is worth exactly what the lab's closed list pays it: one line, its verdict, and your respect for having checked.

Good luck at the windows. Keep the notebook closer than the luck.

Appendices

Appendix A: The closed list

Headnote — 2026-11-05 rebuild. The counts and figures that stood in this appendix at first draft were removed in the 2026-11-05 rebuild because they were measured on a research corpus the lab may not publish. The verdicts stand as the lab's record. The lab's publishable record is its own settled product record — the filed cards of November 2–4, 2026 at first publication — and rebuilt figures elsewhere in this book are sourced to that record.

This appendix is the book's signature feature, and it is deliberately plain. Every method, rule, variant, screen, and feature below was tested by the lab under a rule written down in advance, on a stated population, against a stated benchmark. Each one failed its gates, was reversed on a fuller count, was declined, or was removed. None of them is in the product. None of them is in the chapters as advice.

The format is one line per item: what it claimed, and the verdict the evidence returned. No count is printed in this appendix in this rebuild.

Two reading notes. First, "not supported" is the lab's standard verdict language: the idea was tested as specified and did not clear the gates set before the test ran — paired return against the benchmark, head-to-head duel, and win rate, judged on the primary population with a control population reported alongside. Screens are diagnostic instruments, not pick rules; their lines record whether they separated the population they

were built to flag, and a separating screen is still not an adoption. Second, the list is complete for the sources it is compiled from, not for the lab's whole history: earlier closed items that live only in test reports outside this compilation are not reconstructed here from memory.

Method	What it claimed	Verdict	Races
FAIR threshold family — F1 (favorite only)	A probability threshold on the favorite finds a clean-pass betting subset	Not a clean pass — reversed family	—
FAIR threshold family — U-FLAT	Threshold-selected flat bets beat the controls	Not a clean pass — reversed family	—
FAIR threshold family — U-DUTCH	Threshold-selected dutched bets beat the controls	Not a clean pass — reversed family	—
FAIR threshold family — W2_rank1	Threshold-selected rank-1 bets beat the controls	Not a clean pass — reversed family	—
Within-four re-ranking (R-RANK, Family 4 form)	Re-ranking inside the morning-line top four by median speed and weight beats the four's plain book by the gated margin	Not supported — the margin fell short of its gate	—
§26 consensus as a wagering method	The consensus pick rule's paired edge is proven economics	Annotation only — economics gate failed; interval crossed zero; edge concentrated in one price band and one calendar half	—

Method	What it claimed	Verdict	Races
LOGIT_ALL (locked logistic scorer)	A locked all-factor logistic model beats the favorite on ranking margin and profit	Failed the ranking-margin and profit legs on both populations, by wider margins on the primary; locked paper record only, never refit	—
R-PEI — pace efficiency ranking (lpf/epf)	Ranking a field by late pace retained per unit of early pace beats the morning-line order	Not supported	—
R-CPE — pressure-exposure ranking ($\delta = 5$)	When the favorite faces several rivals with matching early speed, a stalker or closer beats it	Not supported	—
R-PEI×CPE — pace efficiency gated by pressure	The pressure gate rescues pace efficiency as a ranking rule	Not supported	—
R-LEAD — field-relative separation ranking	The horse that leads the most measures by clear margins beats the morning-line order	Not supported — interval crossing zero, duel lost	—
V1 screen — lone speed, weak late	A favorite alone on the early lead with weak late pace is a false favorite	Diagnostic — does not separate; flagged favorites won often	—
V2 screen — closer leaves too much	A deep-closing favorite in a large	Diagnostic only — separates win rate on both	—

Method	What it claimed	Verdict	Races
	field, weak early, is vulnerable	populations, return on the primary only; no adoption from a screen	
R-CV — consistency ranking	The most consistent starter (lowest variation across its recent figures) beats the morning-line order	Not supported	—
R-TREND — improvement ranking	The most improving starter beats the morning-line order	Not supported — interval entirely negative, duel lost	—
R-CAP — class-adjusted ranking	The starter whose figures were earned furthest above the par of its races beats the morning-line order	Not supported	—
R-PEI-BRIS — pace efficiency, BRIS-exact form (LP/E2)	The corpus result reverses when pace efficiency is measured exactly as LP divided by E2 on product cards	Not supported	—
V3 screen — peak outlier	A favorite whose best recent figure sits far above its own range is vulnerable	Does not separate — inverted on the primary population	—
V4 screen — surface exclusivity	A favorite whose recent figures were all earned on	Separating diagnostic on both populations	—

Method	What it claimed	Verdict	Races
	a different surface is vulnerable	— recorded as a screen only; no adoption follows from a screen	
One-hit wonders as a selection pool	A starter outside the morning-line top four that tops one factor list carries usable signal	Measurement only — nothing adopted; the group won rarely and returned less than the leading group	—
Whole-field higher bar HB3	An outsider may take the selection when its consensus total clears the in-four winner by 3 or more points	Not supported — the booked outsiders won rarely and returned badly	—
Whole-field higher bar HB2 (descriptive)	The same construction at a 2-point bar	Not supported (descriptive arm)	—
Whole-field higher bar HB5 (descriptive)	The same construction at a 5-point bar	Not supported — raising the bar made it worse (descriptive arm)	—
B3-ML — three-way rank blend, morning line	Summing a starter's ranks on the raw figure, the adjusted figure, and the morning line beats the plain morning-line order	Not supported	—
LEX — lexicographic sort (raw, then adjusted, then market)	Sorting by raw rank, then adjusted rank, then market rank beats the anchor	Not supported — the weakest arm in its test	—

Method	What it claimed	Verdict	Races
B3-POST — three-way rank blend, final odds	The blend with closing odds in place of the morning line	Oracle ceiling only — uses information not knowable before the race; never a filing candidate	
G-Q — rank-blended figure, quantile form (closing odds and morning-line forms)	Rebuilding each figure from a within-race rank blend of raw, adjusted, and market repairs the figure	Not supported on both populations and both market forms	—
G-N ML — rank-blended figure, nudge form, morning line	The pre-race (morning-line) nudge form of the blended figure beats the current adjusted figure	Not supported — failed at the primary gate; a complement pass cannot rescue a first-gate failure	—
G-N POST — rank-blended figure, nudge form, closing odds	The closing-odds nudge form beats the current adjusted figure	Confirmed — was held, released and adopted 2026-11-05 with the Classic exhibit re-graded	—
Breakout-exception variant, T = 0.5	Charge a race its own variant sample when it diverges from the pooled day variant by at least the threshold	Not supported	—
Breakout-exception variant, T = 1.0	The same construction at a 1.0-point threshold	Not supported	—
Breakout-exception variant, T = 1.5	The same construction at a	Not supported — win rate below	—

Method	What it claimed	Verdict	Races
	1.5-point threshold	the production figure	
Breakout-exception variant, noise-scaled gate (secondary)	Scale the breakout threshold by the race sample's own noise	Not supported	—
A1 — band-split variant	Split the pooled day variant by sprint and route bands	Not supported — win rate below production	—
A2 — race-local shrinkage variant	Shrink each race's own variant sample toward the day variant	Not supported — better ordering, worse prices	—
Variant damping, DIST9 cell, k = 0.75	At nine furlongs and beyond, applying 75% of the variant improves the figure	Supported on the discovery cell, then not confirmed on fresh races; the thread closed as cell noise	—
Variant damping, DIST9 cell, k = 0.50 / 0.25 / 0.00	Deeper damping at nine furlongs and beyond	Not supported at every deeper level	—
Variant damping, graded-stakes purse proxy, all k	Damping the variant in high-purse stakes company	Not supported at any level — fewer winners at better prices; duel negative throughout	—
Variant damping, full population, all k	Damping the variant everywhere	Not supported at any level	—
§14D F2 — elite-connections edge	A starter whose trainer or jockey is winning at 20% or more qualifies	Removed in Protocol v3.17 — its evidence reversed on the corrected count; a	—

Method	What it claimed	Verdict	Races
	its pick under a lighter gate	filing that cites it is a defect	
Selection-keyed trifecta	Keying trifectas on the selection beats the favorite key	Not supported — tested in the verticals framework and not adopted	—
Straight superfecta (selection over 2-3-4 in exact order)	Keying a straight superfecta on the selection beats the favorite key	Not supported — tested in the verticals framework and not adopted	—
Whole-field consensus as the selection source	The whole field's consensus totals, not just the top four's, should name the selection	Not supported as a replacement for the top-four rule; it remains a display	—
Consensus points as a ranking or a probability	The consensus construction ranks fields and prices them	Not supported as a ranking — its points cannot price; adopted only as a pick rule	—
Rank-sum consensus variant	Averaging factor ranks instead of awarding points	Trailed the points form in the variant sweep and was not adopted	—
FAIR as a consensus column	The blend probability, added as a scored consensus factor, improves the pick rule	Earned no seat — split halves	—
LATE as a consensus column	Late pace, added as a scored consensus factor, improves the pick rule	Earned no seat	—

Method	What it claimed	Verdict	Races
MID/TURN as a consensus column	The turn (E2) pace figure, added as a scored consensus factor, improves the pick rule	Earned no seat	—
CONNECTIONS as a consensus column (standing alone)	A connections column improves the pick rule on its own	Not adopted — it passed only beside the draw column and failed alone in every form	—
MWETDC — the sweep's best consensus arm	Dropping speed and adding mid, draw, and connections together improves the pick rule	Not adopted — a three-change arm; the adoption rests on the single-change evidence only	—
Top-2, top-3, and top-5 consensus pools (every bonus form)	A shallower or deeper consensus pool than the top four improves the pick rule	Failed — the shallow pools won their duels and lost money	—
P2 — general off-blend edge requirement	Require a general edge beyond the favorite-over-blend-top gate	Declined — the live arm reversed its duel	—
P3 — age-conditional gate form	An age-conditional form of the favorite-over-blend-top gate	Declined — a dominated subset of the gate it conditioned	—
Whole-surface turf exclusion	Exclude every turf race from selections	Not adopted — the turf-versus-dirt gap fell under the materiality bar and failed stability; two turf cells were excluded instead	—

Method	What it claimed	Verdict	Races
§20-3 — line-store stage-1 features	Computing the blend's starts, days, and wins from the full line store beats the printed past-performance window	Removed in Protocol v3.6 — on corrected data, its gates failed	—
Tier B — adjusted-figure substitute (stage 1)	Substituting the lab's adjusted figure for the published figure inside the blend's first stage	Failed every gate in both its tests; stays closed	—
Layoff × class and days-since-last-win features	Layoff interacted with class, and days since last win, improve the blend's first stage	Declined — all candidate features failed the gates; the literature claims did not reproduce	—
No-morning-line leg removal	Removing filed legs on horses carrying no morning line improves the book	Failed its live split; not adopted	—
Per-track blend calibration (live-era form)	Fitting the blend's market calibration separately per track improves it on live data	Failed its holdout gates in both directions on the live fit; the global form stands	—
§2 no-qualifier path (slot 1)	A marginal late leader with a tightened second edge may file the first slot without the standard qualifier	Remains closed for slot 1	—

The list ends where the sources end. A method that is not in this appendix was either adopted — and the chapters say so, with its record — or it was not tested in the sources this appendix is compiled from. Nothing was left off this list because its numbers were awkward. Several of these ideas were the lab's own, two of them were adopted before fuller evidence reversed them, and one of them — the distance-damping cell — passed its first test outright before fresh races closed it. That is what the standard costs, and it is the point of keeping the list in the book.

Appendix B: Glossary for the casual player

Plain definitions, in the sense this book uses the words. Racing terms first, then the lab's testing terms.

Backtest. A test of a rule on races that had already been run when the rule was written or refined. Necessary, and exploratory: a backtest shows a rule is worth watching, not that it has performed live.

Breakage. The rounding down of payoffs to the nearest dime or nickel, kept by the track. A small, steady charge on top of the takeout.

Duel. The head-to-head count between two methods on the races where they name different horses: how many of those races each method's pick won. Reported as a net — wins for one side minus wins for the other. A method that cannot win its duels is not beating the method it disagrees with.

Morning line. The odds printed in the program before betting opens, set by one track handicapper. It is a forecast of how the crowd will bet, not a measurement of the horses and not the price you will get.

Overlay. A horse whose price is longer than its true chance of winning justifies — the crowd is offering more than the risk is worth. The opposite, an underlay, is the best horse at a price too short to pay for its risk.

Paired difference. The average per-race difference in return between a method and its benchmark, computed race by race on the same races and expressed in points of return on the stake. Pairing removes the races both methods agreed on from the comparison's noise and is the lab's primary money measure.

Par. The typical figure for a kind of race — a track, distance, surface, and class — against which an individual performance is measured. A horse that runs to par ran the race that kind of race usually produces. Class pars are built by collecting the figures of many such races; an expanding par uses only races run before the race being judged.

Past performances (**PPs**). The printed record of a horse's previous races: dates, tracks, distances, surfaces, running positions, margins, times, and figures. The sheet every other measurement in this book starts from.

Post time. The scheduled start of a race. "At post" odds and prices are the final ones, set when betting closes as the race starts.

Pre-registration. Writing a test's rule, population, benchmark, and passing standard down in a fixed form before any outcome is computed. It prevents the quiet editing — a shifted cutoff, a dropped track — that turns a losing idea into a winning story after the fact.

Prospective. A test or record accumulated after a rule was frozen, on races nobody had seen when the rule was written. The only evidence that measures performance rather than fit to old races.

ROI (return on investment). What came back, minus what went in, divided by what went in, stated as a percentage. In this book it is always computed on $2 win bets unless a passage says otherwise: staking $2,000 and returning $1,600 is an ROI of –20%.

Scratch. A horse entered in a race that does not start. Scratched horses are removed from every count in this book before the count is made; a method never gets credit or blame for a horse that did not run.

Takeout. The fixed share of a betting pool the track removes before paying winners — roughly 15 to 17 cents per dollar in a typical North American win pool, and more in most exotic pools. It is the fee every method in this book has to beat before it can make money.

Variant. The adjustment that turns a raw time into a comparable figure: how fast or slow the track itself was playing on the day, estimated from the day's races and charged against each performance. Two horses with the same raw time on different days can earn different adjusted figures because the variant differed.

Win rate. The share of a method's picks that won, as a percentage. It answers "how often was it right?" and nothing about money by itself.

Two more terms the chapters use for kinds of races. A **maiden** race is for horses that have never won; a maiden is such a horse. A horse is a **favorite** when the crowd sends it off at the shortest price in its race — the benchmark every test in this book had to beat.

Appendix C: Where to get past performances and charts

Every test in this book starts from two documents: the past performances, printed before the race, and the chart, the official record published after it. This appendix lists where a casual player gets both. Free sources come first, because the lab's own daily product is built from free sources plus Brisnet past performances, and a reader who wants to repeat a test in Chapter 19 needs the free route to work. Paid sources are named by vendor where the lab's research verified them, and otherwise by type. Prices appear only where the research behind this appendix found them stated plainly on a vendor's own pages, on November 4, 2026; prices change, and the vendor's page controls. Nothing in this appendix is an affiliate listing, and a listing is not a guarantee that any given file will work with any given tool.

Free past performances

Trainer-directory routes. Sites that aggregate publicly posted past-performance files organize them by trainer, by sire, and by track. The lab's morning pull works from its own trainer

directory — a page of links to past-performance summaries published for individual trainers — and sweeps the aggregator directories as well: the trainer and sire directories at thefreepps.com, which curates links to publicly posted Brisnet and Equibase files and produces no files of its own, and the free past-performance page at Horse Racing Nation, which posts downloadable Brisnet files for major races — the Kentucky Derby and Oaks, the Preakness, the Belmont Stakes, and the Breeders' Cup among them — courtesy of Brisnet. These routes give a player full past-performance documents, race by race, at no cost, for the races the sources cover.

Equibase entries. Equibase, racing's official data collector, publishes entries for every North American card free: the horses, trainers, jockeys, morning lines, and program numbers for each race, before it runs. Entries are not past performances — they carry no running lines or pace figures — and the lab uses them for exactly what they are: an index. When a race's past performances cannot be reached directly, the entries page names each horse's trainer, and the trainer name leads back to the directory route above. Equibase also offers a free abbreviated product, Pocket PPs, with each horse's past four races.

Other free offers. Daily Racing Form posts one featured race's past performances free each day (the Race of the Day) and free international past performances for several overseas jurisdictions. Brisnet posts free PDF past performances for British, Irish, Japanese, and other international tracks. Betting sites are a distribution channel rather than a vendor: a wagering

account that bets a card commonly receives that card's Brisnet past performances at no charge, and individual tracks sometimes post free files for their own cards.

Free charts

The chart is the official record of a finished race: the running order, margins, final time, payoffs, and exotic results. Equibase publishes charts free, and they are the lab's source of record for settling every test and every published card: when a chart exists, the chart governs, and a settlement made from any lesser source is corrected when the chart appears. Daily Racing Form also posts free PDF charts for the most recent seven days. A reader running the kitchen-table test in Chapter 19 needs nothing more than these: a rule written down in advance, the past performances from the free routes above, and the charts to score it.

Paid past performances, by type

For a player who wants a full card in one purchase, the major vendors sell downloadable past-performance documents directly. **Brisnet** sells the Ultimate Past Performances in PDF — stated at $3 per card, with a condensed tier stated at $1 per card — and data files from $1 per race card; the lab's own product is built from these documents, gathered through the free routes above wherever they reach. **Daily Racing Form** sells its Classic past performances in PDF, an interactive edition whose print view saves as a PDF, and bundled editions ($4.95 for one track, $9.45 for three, through the Equibase store). **Equibase** sells its own Premium and Basic past-performance programs ($2.00 and $1.50 per card as stated) and resells other vendors' products through the same store. **TrackMaster**, an

Equibase company, sells Platinum Past Performances in PDF ($2.50 per card as stated) and race-card data files in XML. A fifth major product, sold through Daily Racing Form, presents the same document with its own pace and speed figures.

Beyond the major vendors, two other types of paid product exist, and the distinction matters for testing. **Figure sheets** — published by figure-making services, at prices that run to $25 per card for the established sheets — present each horse's figures as graphs and carry no running lines; they are a measurement, not a past-performance document. **Data files** — comma-delimited or XML files sold by data vendors and by the major vendors' data arms — carry the same content as the document in machine-readable form, for players who run their own software; one established data service sells its files by monthly subscription ($89 to $119 per month as stated) with free software that renders them into readable past performances.

The working rule the lab follows, offered as the appendix's closing note: start with the free sources. They cover the races most casual players actually play, they are the same sources this book's own product is built from, and a test that cannot be run on free records cannot be checked by the reader who paid for this book.