CONTENTS

FOR DTC FOUNDERS AND OPERATORS · A FIELD GUIDE

THE NOISE FLOOR

Luck, variance and other people's playbooks, and how to tell a signal from a bad week.

ANDREW LAUCHNER Author of The Second Order and The Whole Machine

September 2026 · 14 chapters · About 65 minutes

A note before you start

Every Monday, somebody explains last week. Revenue fell 9% because the new creative is tired. The welcome flow's conversion rate jumped because of the subject line. Repeat rate dipped because of the shipping delay. Most weeks, most of those stories are about noise.

47%	6.5 pts
of market pioneers failed, in a historical study of about 500 brands in 50 categories. Earlier studies that left out nonsurvivors had missed them (Golder and Tellis, 1993)	a year: the gap between the market's 17.9% and the 11.4% earned after costs by the most active fifth of 66,465 households, 1991 to 1996 (Barber and Odean, 2000; gap Derived)

Those two numbers are the guide in miniature. The first is what happens when you learn only from the companies still standing: the lesson you draw can be the opposite of the truth. The second is what happens when you act on every movement: the most active traders did worst. Operators make both mistakes every week. We copy the brand that won and never meet the ten that did the same thing and vanished. We rewrite the flow, the ad and the offer because a number moved, when the number would have moved anyway.

> **A number that moves isn't news. A number that moves more than it usually does might be.**

The noise floor is a term from audio engineering: the background hiss of a system, below which you can't hear a real signal. Every metric on your scorecard has one. This guide shows you how to measure it, how to hear what rises above it, and how to make decisions before you know how they'll turn out: from base rates, with kill criteria set in advance, and with a written record you can learn from later.

It doesn't teach A/B testing; The Honest Test does. It assumes you know your contribution margin and cohorts, which The Whole Machine covers.

How to read it

Start with The Judgment Audit. Your lowest checks name the chapters to read first. Or follow a path:

- **Founder:** Most Movement Is Noise, Every Winner Falls Back, Other People's Playbooks, Kill Criteria.
- **Running the weekly meeting:** The Process Behavior Chart, Leave a Steady Process Alone, The Signal Scorecard.
- **Planning a launch:** Start From the Base Rate, Kill Criteria, The Premortem.
- **Analyst:** part one, then For Your Analyst.

Three tools run in the page. Nothing you type leaves your browser.

What's proven and what isn't

- **Filed:** numbers from securities filings and SEC actions, here Casper's, Allbirds' and Sunbeam's.
- **Published:** peer-reviewed research and books and papers by the researchers themselves, with the date.
- **Reported:** what a named outlet, company or practitioner published.
- **Derived:** my arithmetic, using a standard formula, with the inputs shown.

Examples that open with Say or Picture use made-up round numbers. Every source is listed in Appendix C.

Andrew Lauchner

Scottsdale, Arizona

FRONT

TEN POSITIONS

What this guide argues, and what would prove each claim wrong.

A position says what would prove it wrong. Test each on your own numbers.

01 **Most week-to-week movement in revenue, conversion, repeat rate and email metrics is noise.**

Wrong if, on a process behavior chart of the last year, more than a handful of weeks fall outside the limits.

Most Movement Is Noise

02 **A weekly number needs limits drawn from its own history before anyone explains it.**

Wrong if the explanations your team gave for last quarter's weekly swings predicted the next week's direction better than a coin.

The Process Behavior Chart

03 **Adjusting a steady process in response to noise makes it worse.**

Wrong if your most frequently rewritten flow or campaign type outperforms the ones you left alone.

Leave a Steady Process Alone

04 **The best campaign, SKU or channel of any period will usually do worse next period.**

Wrong if last quarter's top five campaigns, resent to similar audiences, did as well the second time.

Every Winner Falls Back

05 **Skill shows up as persistence. A result that doesn't repeat wasn't skill.**

Wrong if the channels, creatives or agencies that ranked highest in one quarter don't rank higher than average in the next.

Skill, Luck and Persistence

06 **Case studies and founder threads describe survivors, written after the outcome was known.**

Wrong if you can find the brands that did the same thing and failed, and they're rare.

Other People's Playbooks

---

07 **Forecast from what happened to projects like yours before you forecast from your plan.**

Wrong if your last five launches hit their plans without a reference class.

Start From the Base Rate

---

08 **The time to decide when to stop is before you start.**

Wrong if the projects you stopped late were stopped for reasons nobody could have written down at launch.

Kill Criteria

---

09 **Imagining a project has already failed surfaces risks a plan review misses.**

Wrong if a premortem on your next launch finds nothing the plan review didn't.

The Premortem

---

10 **Judge decisions on the information and process you had, not on how they turned out.**

Wrong if your team can say, without a written record, what it expected from its last ten big decisions.

The Decision Journal

FRONT

THE NOISE FLOOR ON ONE PAGE

Eight questions operators ask every month. Each has a trap that makes the obvious answer wrong, and a tool that fixes it.

Find the question you asked last week. Read across.

THE QUESTION	THE TRAP	THE FIX	CHAPTER
Did revenue really drop this week?	Explaining noise	A process behavior chart with limits from your last 12 weeks	3
Should we rewrite the flow?	Tampering with a steady process	Change only on a signal or a plan	4
Is this campaign a winner?	Regression to the mean	Shrink the result toward your average	5
Is this channel, agency or hire good?	Calling luck skill	Check whether rank persists from one period to the next	6
Should we copy what they did?	The halo effect and survivors	Ask the evidence questions first	7
Will this launch hit plan?	The inside view	Forecast from a reference class, then adjust	9
When do we stop?	Deciding under sunk cost	Kill criteria written before launch	10
Was that a good call?	Judging by the outcome	A decision journal written at the time	12

The first four rows are about reading what already happened. The last four are about deciding before you know. Both halves rest on the same idea: an outcome is a draw from a

range of possible outcomes, and you only ever see one draw. The job is to estimate the range, act on the range, and learn from the draws without treating each one as a verdict.

> **You only see one draw. Decide on the range.**

Do this

- **Write down the last three weekly-meeting explanations** your team gave for a number that moved. Keep the list. By the end of chapter 3 you'll know which of them explained a signal and which explained noise.

START HERE · CHAPTER 1

THE JUDGMENT AUDIT

Twelve checks on whether your team reacts to signals or to noise, and whether it decides before it knows or after. About forty minutes with your scorecard, your flow history and your last three launch plans.

This audit isn't about whether your numbers are good. It's about whether the way you read them, and the way you decide, can tell a real change from a normal week, and a good decision from a lucky one.

Open your weekly scorecard, the change log for your email and SMS flows, your last three launch or campaign plans, and your notes from the last few weekly meetings. Score each

check 0 to 2: 0 if it failed or nobody can answer it, 1 if partly true, 2 if clean. “We usually talk about it” scores 0. The check is whether it’s written down.

> **If it isn’t written down before the result comes in, it didn’t happen.**

The twelve checks

01 **Weekly numbers have limits · 5 min**

Look at: Your weekly scorecard.
Good: Revenue, orders, conversion rate and the main retention numbers each show a central line and upper and lower limits computed from their own recent history.
Cost if wrong: Every wiggle gets a story and an owner, and the one change that matters looks like the rest.
Read next: The Process Behavior Chart

02 **Explanations wait for a signal · 3 min**

Look at: Notes from your last four weekly meetings.
Good: Numbers inside their limits were noted and left alone. Explanations and actions were reserved for numbers outside them.
Cost if wrong: The team learns causes that aren’t real and repeats them in next quarter’s plan.
Read next: Most Movement Is Noise

03 **Small denominators are labeled · 3 min**

Look at: Any rate on the scorecard built on fewer than about 1,000 people or 100 events a week: a segment’s click rate, a new SKU’s sales, a small cohort’s repeat rate.
Good: Each shows its denominator, and nobody acts on a move that chance alone could produce.
Cost if wrong: You kill good SKUs and promote lucky ones.
Read next: Most Movement Is Noise

04 **Open rate isn’t a decision metric · 2 min**

Look at: How email subject lines and sends are judged.
Good: On clicks, orders or revenue per recipient, not opens.
Cost if wrong: You optimize a number that Apple Mail inflates for reasons unrelated to your subject line.
Read next: Most Movement Is Noise

05 **Flows change on a plan or a signal · 4 min**

Look at: The change log for your five highest-revenue flows over the last six months.
Good: Each change followed a signal on the chart, or was a planned test with a holdout. No flow was rewritten more than twice without one.
Cost if wrong: Tampering adds variation, and nobody can say which version worked.
Read next: Leave a Steady Process Alone

---

06 **Winners are shrunk before they're scaled · 3 min**

Look at: Your last three "winning" campaigns, creatives or SKUs that got more budget or inventory.
Good: Someone estimated how much of the result would hold, and planned on that figure rather than the headline.
Cost if wrong: You buy inventory and budget for a result that was partly luck.
Read next: Every Winner Falls Back

---

07 **Judgments of channels and people use two periods · 4 min**

Look at: How you last judged an agency, a channel or a team member on results.
Good: You compared at least two separate periods, and the ranking held.
Cost if wrong: You fire the unlucky and promote the lucky, then wonder why the new star fades.
Read next: Skill, Luck and Persistence

---

08 **Copied practices have a written reason · 3 min**

Look at: The last tactic you adopted because another brand did it.
Good: Someone wrote down why it worked there, whether brands that tried it and failed exist, and why it should work here.
Cost if wrong: You import a survivor's story and none of its conditions.
Read next: Other People's Playbooks

---

09 **Forecasts start from a reference class · 4 min**

Look at: Your last three launch forecasts: a product, a channel, a market.
Good: Each began with what similar launches, yours or others', actually did, and then adjusted for the specifics.
Cost if wrong: Plans anchor on the best case, and inventory and hiring follow the plan.
Read next: Start From the Base Rate

---

10 **Live bets have kill criteria · 3 min**

Look at: Every initiative now running that costs real money or time.
Good: Each has a written condition and a date at which it stops, decided before it started.

Cost if wrong: Projects run on sunk cost and hope for a quarter longer than they should.
Read next: Kill Criteria

11 **A premortem before the last big launch · 2 min**

Look at: Your last major launch.
Good: Before it went live, the team spent twenty minutes imagining it had failed and writing down why, and the plan changed because of it.
Cost if wrong: The risks everyone privately saw stay private.
Read next: The Premortem

12 **A decision journal · 3 min**

Look at: Where big decisions are recorded.
Good: Each has a dated entry written at the time: what you expected, how confident you were, and what would change your mind.
Cost if wrong: Hindsight rewrites every decision, and you learn from outcomes instead of from reasoning.
Read next: The Decision Journal

Score as you go; your band appears when all twelve are in.

RUN YOUR NUMBERS

Score the twelve checks

0: failed, or nobody can answer it. 1: partly true. 2: clean. Scores stay in this browser.

01 Weekly numbers have limits 5 min	0	1	2
02 Explanations wait for a signal 3 min	0	1	2
03 Small denominators are labeled 3 min	0	1	2
04 Open rate isn’t a decision metric 2 min	0	1	2

05 Flows change on a plan or a signal 4 min	0	1	2
06 Winners are shrunk before they're scaled 3 min	0	1	2
07 Judgments of channels and people use two periods 4 min	0	1	2
08 Copied practices have a written reason 3 min	0	1	2
09 Forecasts start from a reference class 4 min	0	1	2
10 Live bets have kill criteria 3 min	0	1	2
11 A premortem before the last big launch 2 min	0	1	2
12 A decision journal 3 min	0	1	2

0	0 of 12
of 24 points	checks scored

Score all twelve to read your band. 12 to go.

RESET SCORES

Read your score

SCORE	WHAT IT MEANS	READ NEXT
20–24	Your team reads signals, not noise, and decides before it knows. Keep the records going and review them each quarter.	The Signal Scorecard, then The Decision Journal
14–19	You read the numbers well in places and react to noise in others. Fix the zeros first.	The chapter linked from your lowest check, then The Process Behavior Chart
8–13	Most of what gets explained on Monday is noise, and most big bets are judged by how they turned out.	Part one, starting at Most Movement Is Noise
0–7	Stop explaining weekly moves for a month. Put limits on five numbers and write kill criteria for every live bet.	The Process Behavior Chart, then The First Thirty Days

If checks 9 to 12 averaged lower than checks 1 to 8, start with part three. Reading numbers is something your team already practices every Monday. Writing things down before the result arrives is a new habit, and it’s the cheaper one to build.

PART ONE · READING THE WEEK · CHAPTER 2

MOST MOVEMENT IS NOISE

Before anyone explains a number, ask how much it moves when nothing has changed. The answer is usually more than the move you're explaining.

Every weekly number is the sum of thousands of small causes: who happened to be paid that Friday, the weather in Texas, what the ad auction did on Tuesday, which day the email went out, a creator's post you never saw. Those causes don't stop. They make every number wobble, every week, with no story behind the wobble.

Chance alone, before anything else

Start with the smallest source of wobble: pure chance. Say a brand ships 10,000 orders a month, about 2,300 a week. Even if nothing about the store, the traffic or the customers changed, the weekly count would bounce around 2,300 the way coin flips bounce around half heads. The 95% range from chance alone is about 94 orders either way, or 4%. Two weeks compared with each other can differ by almost 6% by chance alone DERIVED .

That's the floor, and it's the best case: a big count and a simple metric. Smaller denominators wobble far more.

NUMBER	TYPICAL WEEK	CHANCE ALONE, EITHER WAY	AS A SHARE OF THE NUMBER
Weekly orders	2,300 orders	±94 orders	±4%
Site conversion rate	2.5% of 90,000 sessions	±0.10 pts	±4%
Email click rate	1.5% of 20,000 recipients	±0.17 pts	±11%

NUMBER	TYPICAL WEEK	CHANCE ALONE, EITHER WAY	AS A SHARE OF THE NUMBER
Repeat rate, one month's cohort	25% of 400 customers	±4.2 pts	±17%
One SKU's sales	40 units	±12 units	±31%
A campaign's order rate	0.15% of 12,000 recipients	±0.07 pts	±46%

DERIVED 95% ranges: 1.96 × √(p(1 − p) / n) for rates and 1.96 × √n for counts. This is sampling chance only. Real weeks vary more.

Read the last two rows again. A campaign to 12,000 people that normally produces 18 orders can produce 26 or 10 with nothing different about it. A new SKU selling 40 units a week can sell 28 or 52. Those are the numbers most teams make their fastest decisions on.

> **The smaller the denominator, the louder the noise. The loudest numbers get the fastest decisions.**

Real weeks are noisier than chance

The table understates it, because the thousands of small causes add their own variation on top of sampling chance. Walter Shewhart of Bell Laboratories, who invented the control chart, called them chance causes, and the rare big ones assignable causes; W. Edwards Deming later called the second kind special causes **REPORTED** . A good week and a bad week from the same steady system are both made of chance causes. Nobody can find "the reason" for one of them, because there isn't one.

You can't compute that total wobble from a formula. You measure it from your own history, which is what the chart in the next chapter does. How wide it is depends on the business. In the example in the next chapter, the limits sit about 20% either side of average weekly revenue, and that store has nothing wrong with it.

Email numbers have extra noise

Open rate is the noisiest number in the building, and not only because of chance. Apple's Mail Privacy Protection downloads remote content "in the background by default," whether or not the reader engages with the email, and loading that content is what makes an open register. Apple says it stops senders from learning when or how many times a message was opened REPORTED . So a change in opens can come from how many of your recipients use Apple Mail on a given send. Judge subject lines and sends on clicks, orders or revenue per recipient. For deliverability, which does use opens as one input, see The Whole Machine.

Campaign results carry the small-denominator problem too. Most campaigns produce a few dozen orders, and a few dozen orders move by a third on chance alone. One campaign doesn't tell you whether an offer, a subject line or a day of the week works. Twenty campaigns might.

Repeat rate moves slowly and gets read fast

A monthly cohort of 400 first-time buyers can show a 30-day repeat rate of 21% or 29% from the same underlying behavior. Read repeat rate by cohort and over at least three cohorts before you credit a new post-purchase flow or blame a shipping delay. The Second Order covers how to build the cohort view.

Two ways to be wrong

There are two mistakes, and fixing one makes the other more likely. You can treat noise as a signal: explain it, act on it, change something that wasn't broken. Or you can miss a real signal because every week already has a story and the real one looks like the rest. Shewhart's chart was designed to balance the two, so that you react rarely enough to avoid chasing noise and often enough to catch real change. Gut feel guards reliably against neither.

Do this

- **Add the denominator next to every rate on your scorecard.** Anything built on fewer than about 1,000 people or 100 events a week gets a gray label: "small sample, read monthly."
- **Remove open rate from any decision** about subject lines, send times or winners. Keep it on the deliverability page only.

PART ONE · READING THE WEEK · CHAPTER 3

THE PROCESS BEHAVIOR CHART

Two lines drawn from your own last twelve weeks tell you which moves deserve a meeting. The math fits on an index card.

On May 16, 1924, Walter Shewhart sent his superiors at Bell Laboratories a memo proposing a simple chart REPORTED . A century later, the version Donald Wheeler calls the process behavior chart, or XmR chart, is still the best tool for a weekly scorecard. It tells you how much a number moves when nothing has changed, from the number's own history.

Building one

01 **List the last 12 weeks**

Revenue, orders, conversion rate: any number you report weekly. Leave out weeks with a big promotion, or chart promotion weeks on their own.

02 **Take the average**

That's the central line.

03 **Take the moving ranges**

The difference between each week and the week before, ignoring the sign. Twelve weeks give eleven moving ranges. Average them.

04 **Draw the limits**

Average plus and minus 2.66 times the average moving range. Those are the natural process limits. The moving ranges get their own upper limit: 3.268 times the average moving range.

PUBLISHED Wheeler, "What Makes the XmR Chart Work?" Quality Digest, 2012; *Understanding Variation*, 1993. The 2.66 is 3 divided by 1.128, the constant that converts an average moving range into an estimate of the standard deviation, so the limits sit about three standard deviations from the average.

Why the moving range and not the standard deviation of all twelve weeks? Because if the number shifted partway through, the overall standard deviation includes the shift and the limits come out too wide to show it. The moving range looks only at week-to-week changes, which a shift barely touches. It's also why the constant is 2.66 and not 2 or 3: people who swap in a rounder number get limits that are too tight and flag noise, or too loose and miss signals.

The three rules

Wheeler's book gives three rules for spotting a signal REPORTED :

01 **One point outside the limits**

A big, sudden change. Find out what happened that week.

02 **Three of four in a row closer to a limit than to the central line**

A moderate change that has lasted a few weeks.

03

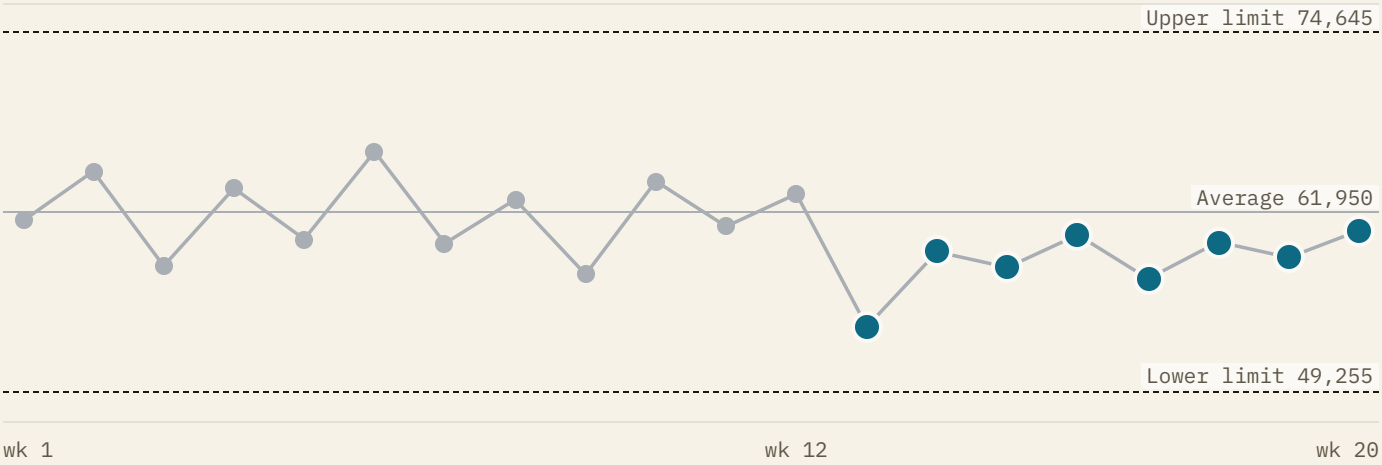
Eight in a row on the same side of the central line

A small, sustained shift. The easiest to miss by eye and often the most important.

For the moving range, use only the first rule: a single week-to-week jump above its upper limit PUBLISHED . Everything else is noise. It gets noted, not explained.

The scary week was noise. The quiet shift after it wasn’t.

Weekly revenue, dollars. Limits from weeks 1 to 12.



Picture a store with these numbers. Week 13 fell 15% from week 12 and got a meeting; it was inside the limits. Weeks 13 to 20 sat below the average eight times in a row, a drop of about 5% that nobody noticed. That was the signal.

That’s the pattern the chart exists to catch. The big drop felt like news and wasn’t. The real change was small, steady and invisible on a week-over-week report, because every one of those weeks, compared with the week before, looked normal.

RUN YOUR NUMBERS

Is this week a signal?

Example numbers. Replace with yours: twelve normal weeks, week 1 the oldest, then the week you’re asking about. Revenue, orders, or a rate in percent all work.

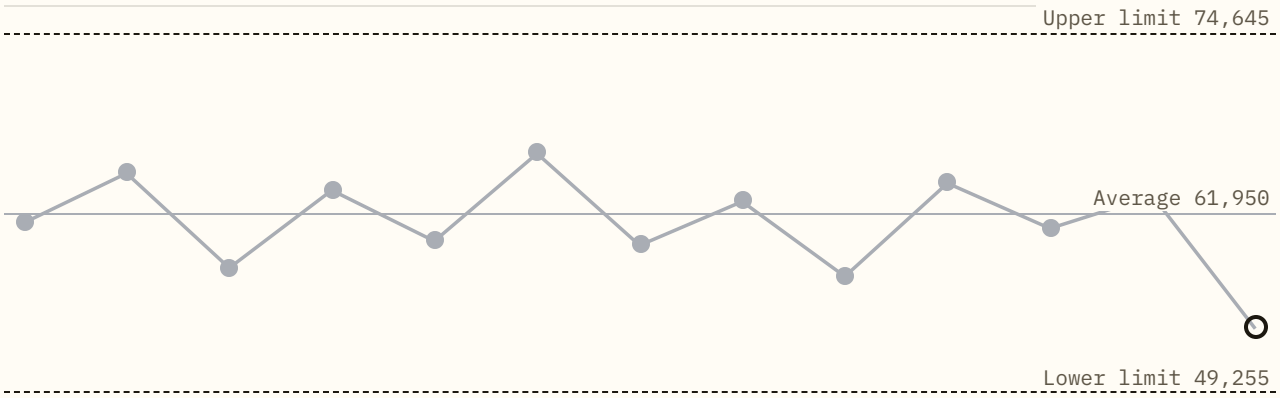
Week 1	Week 2	Week 3	Week 4
61400	64900	58200	63700
Week 5	Week 6	Week 7	Week 8

60100	66300	59800	62900
Week 9	Week 10	Week 11	Week 12
57600	64200	61000	63300

This week, the one you're asking about

53900

61,950	49,255 to 74,645	−13.0%	9,400 of 15,597
average of the 12 weeks: the central line	natural process limits (average ± 2.66 × average moving range)	this week (53,900) against the average; −14.8% against last week	this week's jump from last week, against its limit



Week · This week · Signal · Limit

> **Noise. This week is 13.0% below the average but inside the limits, and no rule fires. Note it and leave it alone. On its own, only a week below 49,255 or above 74,645 would be a signal.**

Limits come from weeks 1 to 12 only. Rules checked: a point outside the limits; three of four in a row closer to a limit than to the average; eight in a row on one side; a week-to-week jump above 3.268 times the average moving range. Leave out promotion weeks, or chart them separately.

With the defaults, this week's $53,900 is 13% below the average and 15% below last week, and it's noise: the limits run from $49,255 to $74,645 and no rule fires. Change this week to 48,000 and the verdict flips to a signal. The rule-of-eight shift in the figure above doesn't

show up here, because the tool judges one new week; for runs, keep charting week after week.

Keeping it honest

- **How many weeks.** Twelve is a workable start. Practitioners who teach the method suggest limits begin to firm up around 17 to 25 points REPORTED . Keep the same limits as you add weeks; recompute only when a signal shows the process has changed, using the weeks since the change.
- **Seasons and promotions.** A sale week is a planned special cause. Chart it against other sale weeks, not against normal ones. If your business is strongly seasonal, chart this year divided by the same week last year instead of raw revenue.
- **Rates and counts both work.** Chart the rate (conversion, repeat, click rate) and its denominator side by side, so a change in traffic doesn't pass for a change in behavior.
- **Mark the changes.** Put a dated note on the chart every time something is launched, repriced or rewritten. A signal near a note is a lead; a signal without one is a question.

Do this

- **Chart five numbers this week:** revenue, orders, site conversion rate, revenue per email recipient and the repeat rate of your most recent full cohort. Put the limits on the scorecard, and read the three rules before anyone offers an explanation in the Monday meeting.

LEAVE A STEADY PROCESS ALONE

Changing a steady process because of a normal week doesn't just waste effort. It adds variation. Investors show the cost in hard numbers.

Deming had a name for adjusting a steady process in response to noise: tampering. He demonstrated it with his funnel experiment, in which people try to hit a target by moving the funnel after each drop, and scatter the drops wider than if they'd left it alone REPORTED .

Why reacting makes it worse

The arithmetic is simple. Suppose each week lands on the true average plus some random error. If you leave the process alone, the weeks scatter with that error's spread. If you "correct" each week by the amount it missed last time, every result now carries two errors: this week's, and the reversal of last week's. The spread of results doubles in variance, about 41% wider in the units you see DERIVED . The team that responds to every dip isn't steering. It's shaking the wheel.

> **Responding to every dip isn't steering. It's shaking the wheel.**

The flow rewrite habit

Picture a welcome flow that converts about 6% of new subscribers, week in, week out, with ordinary wobble. One week it shows 5.1%. Someone rewrites the second email. The next week it's back to 6.2%, and the rewrite gets the credit. A month later another dip, another rewrite. After a year the flow has had eight versions, nobody knows which one was best, and the flow is exactly as good as it was, with more variation and a longer change log.

Each rewrite has costs that don't show on the chart: the hours, the review cycle, the risk of a broken link or a wrong discount code, and the loss of any clean comparison. Worse, the

regression you'll meet in the next chapter makes the habit self-reinforcing. A bad week is usually followed by a better one, so the fix always seems to work.

The overtrading parallel

Brad Barber and Terrance Odean studied 66,465 households with accounts at a large discount broker from 1991 to 1996. The fifth of households that traded most earned 11.4% a year after costs; the market returned 17.9%. The average household turned over 75% of its portfolio a year PUBLISHED . In a follow-up on more than 35,000 households, men traded 45% more than women, and trading cut their net returns by 2.65 percentage points a year, against 1.72 for women PUBLISHED . Their explanation was overconfidence: people who trade most are surest that they know what the next move will be.

A store isn't a portfolio, and a flow rewrite isn't a stock trade. But the mechanism carries over. Each action feels informed, has a small cost, and responds to movement that's mostly noise. Added up, the activity costs more than it earns. The operator who rewrites the flow every month is trading on noise and paying the commission in hours and in lost learning.

When to change something

- **On a signal.** A point outside the limits or a run means something changed. Find out what before you change anything back: a broken discount code, a tracking change or a stock-out is a cause to fix, not a copy problem.
- **On a plan.** If a steady process is steadily not good enough, the whole system needs a deliberate change: a new offer, a new sequence, a new page. Make it once, with a holdout if you can (see The Honest Test), write down what result you expect, and watch the chart for the shift.
- **Never on a normal week.** A normal bad week is the process doing what it does.

Do this

- **Pull the change log for your top five flows** and count the edits in the last six months. For each, mark whether it followed a signal, a plan, or a bad week. Freeze any flow with more than two bad-week edits for the next quarter.

PART TWO · LUCK AND THE WINNERS · CHAPTER 5

EVERY WINNER FALLS BACK

The best campaign, SKU or channel of the quarter was probably good and lucky. The luck doesn't come with it into next quarter.

In 1886 Francis Galton published the heights of 928 adult children of 205 pairs of parents. Tall parents had tall children, but less tall than themselves: on average, a child's height stood about two-thirds as far from the average as the parents' did PUBLISHED . He called it regression towards mediocrity. We call it regression to the mean, and it applies to every number that mixes something real with something random.

The flight instructors

Daniel Kahneman called it "the most satisfying Eureka experience" of his career. He was teaching Israeli Air Force flight instructors that praise works better than punishment for learning a skill, when a senior instructor objected. In his experience, cadets he praised for a clean maneuver usually did worse on the next try, and cadets he screamed at usually did better PUBLISHED .

The instructor's observation was right and his explanation was wrong. He praised after unusually good flights and screamed after unusually bad ones, and unusual flights are followed by more ordinary ones whatever the instructor says. Kahneman's conclusion: because we reward others when they do well and punish them when they do badly, "we are statistically punished for rewarding others and rewarded for punishing them." He then had the instructors toss two coins each at a target behind their backs, without feedback, to show the same pattern appearing with no instruction at all PUBLISHED .

> **Praise follows a lucky week. So does disappointment.**

Where it shows up in a store

- **The winning campaign.** Last quarter's best email, resent to a similar audience, does worse. The copy didn't wear out. Part of the first result was luck.
- **The breakout SKU.** The product that doubled in its second month gets a deep reorder, and month three is merely good. Now there's a warehouse problem.
- **The hot channel.** The ad set, influencer or channel with the best return last month gets the budget, and returns less.
- **The fix that "worked."** The worst flow, page or rep gets attention after its worst month, and improves. Some of that would have happened anyway, which is why a fix needs a comparison, not a before-and-after.

How much falls back

Kahneman's rule: "Whenever the correlation between two scores is imperfect, there will be regression to the mean" PUBLISHED . How far a result falls back depends on how much of the spread between results is real. If your campaigns differ a lot in true quality and each is sent to a big audience, a winner keeps most of its lead. If they're similar in quality and the audiences are small, most of the lead is noise and goes away.

You can estimate the split from your own history. The spread of past campaigns' results is real differences plus chance. Chance you can compute from the audience size. Take it away, and what's left is the real spread. The share of a winner's lead you should expect to keep is

the real spread divided by the real spread plus the chance in that one result DERIVED . The tool does it for you.

RUN YOUR NUMBERS

How much of this win will hold?

Example numbers. Replace with yours. Works for email and SMS campaigns, and for anything else measured as a share of people: orders per recipient, conversion per visitor.

People who got the winner

12000

Orders it produced

30

Average order rate of past campaigns

0.15 %

Spread of past campaigns (standard deviation)

0.05 pts

Typical audience of past campaigns

12000

0.250%	0.200%	50%	24 vs 30
the winner's order rate	what to expect next time	of the lead over average likely to hold	orders to expect next time at the same audience size, against 30 this time and 18 for an average campaign

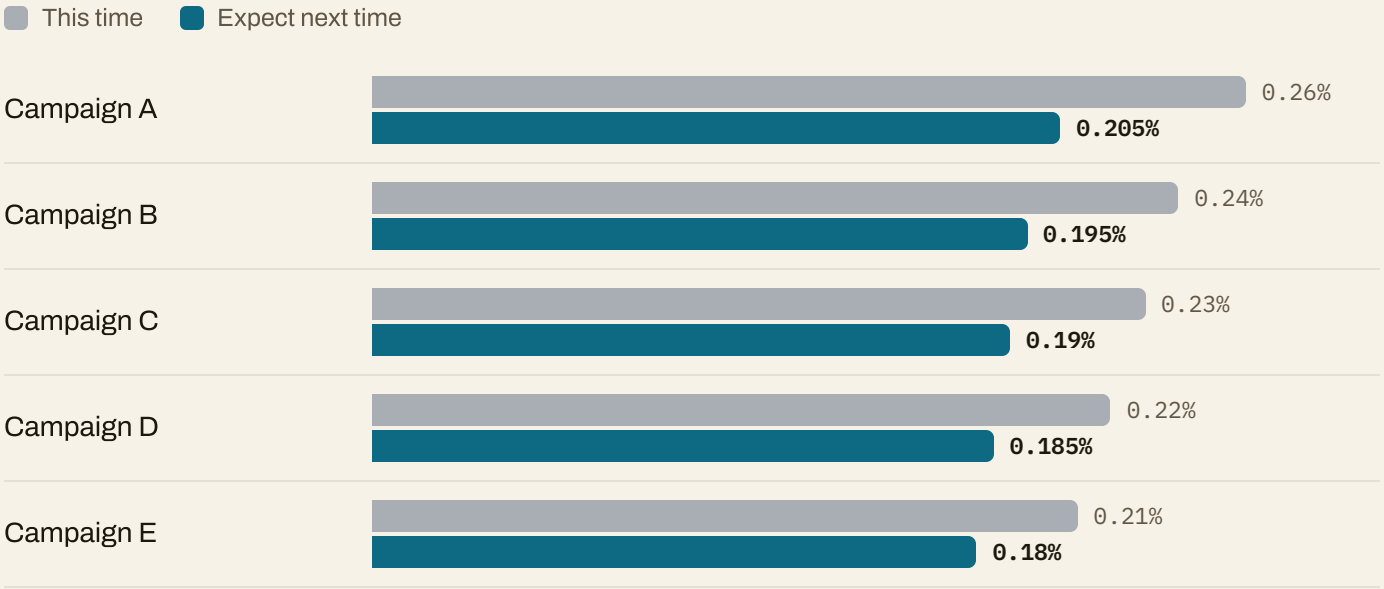
> **Part skill, part luck. Expect about 50% of the lead to hold: plan on 0.200%, not 0.250%. Resend or reuse it, but size inventory and budget to the lower figure.**

Take the spread from the last 15 to 30 comparable campaigns: same channel, similar audience. The estimate assumes the winner was chosen from that same population; if you picked it as the best of many, expect a little less than this.

With the defaults, a campaign to 12,000 people that produced 30 orders, a 0.25% order rate against a usual 0.15%, should be expected to do about 0.20% next time: half its lead is likely real and half was luck. That's 24 orders, not 30. Double the audience of the winner to 24,000 with 60 orders and about two-thirds of the lead holds, because a bigger audience carries less chance.

Last quarter's five best campaigns, and what to expect when you resend them

Order rate per recipient. Picture a store with the defaults above: past average 0.15%, audiences of 12,000.



DERIVED Expected = 0.15% + 0.5 × (observed − 0.15%), where 0.5 is the share of the lead that holds with the default inputs. Bars scaled to 0.30%.

The winners are still winners. They're just closer to the pack than the dashboard says, and the gap between them mostly disappears.

Do this

- **Before any winner gets more budget, inventory or a slot in the calendar,** run it through the tool and plan on the expected figure. For SKUs, use weekly units against the category's weekly units and treat the output as the reorder forecast.

SKILL, LUCK AND PERSISTENCE

Skill repeats and luck doesn't. So the test for skill is whether a ranking holds from one period to the next.

Michael Mauboussin's *The Success Equation* (2012) starts from a simple model: every result is part skill and part luck. You can't see the parts in one result. You can see them across results, because skill carries forward and luck doesn't.

The persistence test

In a 2012 Harvard Business Review article, Mauboussin set out what makes a statistic useful. One test is that it's persistent: the same action produces a similar outcome from one time to the next PUBLISHED . Apply that to anything you're judging on results: channels, creatives, agencies, creators, people.

Say you ran ten creators last quarter and ten again this quarter, the same ten. Rank them by cost per order in each quarter and compare the two rankings. If the top three last quarter are near the top again, results are persistent and the ranking reflects something real. If last quarter's top three are scattered through this quarter's list, the ranking was mostly luck. Spreadsheets do this in one line: the correlation between the two columns of ranks.

- **Correlation near 0:** luck dominates. Don't reallocate on rank. Spread the budget on price and fit, and keep measuring.
- **Around 0.3 to 0.5:** some skill, lots of luck. Tilt budget toward the leaders, gently, and expect the tilt to pay less than the ranking says.
- **Above about 0.7:** the ranking is real. Act on it.

That correlation does double duty. It's also roughly the share of a winner's lead you should expect next period, which is the same shrinkage as chapter 5's tool, measured a different way DERIVED .

> **A ranking that doesn't repeat isn't a ranking of skill.**

The paradox of skill

Mauboussin's paradox of skill: as people get better at an activity, the gaps between the best, the average and the worst become "much narrower" PUBLISHED . When everyone is skilled, luck decides more of the ranking. I'd expect paid social to work this way. When every brand in a category uses the same ad platform, the same creative formats and similar agencies, the difference in skill between them shrinks, and the week-to-week ranking of who's winning becomes noisier. The edge moves to the places competitors can't copy quickly: the product, the offer, the retention program.

Skillful and lucky

Mauboussin put the sports version plainly: a player with a great statistical year was almost always "skillful and lucky" REPORTED . The same holds for a founder who had one great year, or an agency whose case study features its best client. Both parts were there. Only one of them is for hire.

Judging people

Results over one quarter are a poor measure of a person whose results depend on campaigns, markets and budgets they don't control. Judge on two things instead: results over several periods, and the process measures they do control, like whether tests were set up properly, whether launches shipped on time, whether the scorecard was kept. Process measures tend to persist, which is why they're fairer.

Do this

- **Run the persistence test on your biggest allocation decision.** Pick the thing you reallocate most often (channels, ad sets, creators, SKUs on the homepage) and correlate

its ranking in the last two quarters. Write the number at the top of next quarter's budget.

PART TWO · LUCK AND THE WINNERS · CHAPTER 7

OTHER PEOPLE'S PLAYBOOKS

Case studies and founder threads are written after the outcome is known, about the companies that survived. Both facts bend the lesson.

A founder posts the seven things that took the brand to $50 million. Every item is plausible. Some may even be true. But the post is built from two biased samples at once: it describes a winner, and it describes the winner as seen by someone who already knows it won.

The halo effect

Phil Rosenzweig's *The Halo Effect* (2007) takes apart the most popular business research. The halo effect is "the tendency to make specific inferences on the basis of a general impression" PUBLISHED . When a company is doing well, observers describe its strategy as clear, its culture as strong and its leaders as decisive. When the same company does badly, the same observers find confusion, arrogance and drift. His example is Cisco, praised for its strategy in the late 1990s and criticized by many of the same observers when the bubble

burst, though the company hadn't fundamentally changed. The data behind much of the popular research comes from "retrospective interviews, articles from the business press, and business school case studies," all colored by the outcome PUBLISHED .

Rosenzweig's test is one question: if I didn't know how the company was performing, what would I think of its culture, its execution and its customer focus? PUBLISHED

The missing failures

Jerker Denrell showed why learning from successful firms goes wrong even without the halo. Failed firms disappear from view, so the sample you learn from is survivors. His 2003 paper shows how a risky practice with no link to performance across all firms can seem positively related to performance among the survivors PUBLISHED . A bold strategy produces big winners and big losers. Look only at the survivors and it looks like a strategy for winning.

DTC is full of these. Raise big, spend big on brand, open stores, launch a second category. Some brands did those things and thrived. More did them and closed, and nobody writes the thread about them.

First isn't a strategy

Peter Golder and Gerard Tellis tested the belief that pioneers win. Earlier studies had relied on databases of surviving firms. Using historical records for about 500 brands in 50 categories, they found that 47% of market pioneers failed, their mean market share was 10%, and only 11% were still category leaders. Earlier studies had put pioneers' share near 30% and found almost half of them leading. The leaders they found entered on average 13 years after the pioneer PUBLISHED . The pioneers who survived long enough to be studied made first-mover advantage look real.

> **The advice comes from the survivors. The evidence is with the ones who didn't make it.**

Questions before you copy

Jeffrey Pfeffer and Robert Sutton made the case for evidence-based management in Harvard Business Review and in *Hard Facts, Dangerous Half-Truths and Total Nonsense*

(both 2006): as medicine learned to do, managers should decide from the best evidence about what actually works PUBLISHED . The questions below are mine, not theirs. They're what I'd ask before copying anything from another brand:

01 **Did the practice cause the success, or come with it?**

Successful brands do many things. Which of them mattered is rarely known, even to the founder.

02 **Who did the same thing and failed?**

If you can't name any, you haven't looked, not proven they don't exist.

03 **Would I rate this practice the same if the brand were struggling?**

Rosenzweig's question, applied to one tactic.

04 **What conditions made it work there?**

Margin, price point, category, customer, timing, the team. List the ones you don't share.

05 **What would it cost to find out on a small scale?**

A holdout, a single market, one SKU. If the test is cheap, run it instead of arguing.

06 **What's the downside if it's wrong?**

Copying a subject-line style risks little. Copying a store rollout risks the company.

Do this

- **Take the last tactic you copied** and answer the six questions in writing. Then pick one tactic you're considering now and answer them before the meeting where it gets decided.

PRAISED, THEN BLAMED

Three DTC brands that were case studies in both directions, and one celebrated turnaround whose numbers weren't real. What to take from them, and what not to.

These aren't stories about foolish founders. Each brand built something customers liked, and each was written up as a model before it was written up as a warning. The point is how much of both stories was hindsight.

Casper

Casper went public in February 2020 at $12 a share FILED . Its revenue grew from $358 million in 2018 to $439 million in 2019 and $497 million in 2020, while it lost about $90 million in each of those years FILED . In January 2022, Durational Capital Management completed its acquisition of the company for $6.90 a share in cash, and Casper left the New York Stock Exchange FILED . The losses were in the filings the whole time. The story around them changed.

Allbirds

Allbirds priced its IPO in November 2021 at $15 a share FILED . Net revenue rose from $194 million in 2019 to $298 million in 2022, then fell each year to $152 million in 2025 FILED . In 2026 the company disclosed substantial doubt about its ability to continue as a going concern, sold its footwear business's assets, including the brand, to American Exchange Group in a sale that closed on June 9, and renamed itself Smartbird, with a new business acquiring and monetizing the GPU chips used in AI computing. Its 10-Q for the second quarter says the sale proceeds and new financing alleviated the doubt FILED . Sustainability, the wool runner and the founders' story were once cited as the reasons it grew. They were the same things when it shrank.

Glossier

Glossier was valued at $1.8 billion in July 2021. In January 2022 it cut more than 80 corporate jobs, about a third of its corporate staff, many in technology, and in May 2022 its founder Emily Weiss stepped down as chief executive to become executive chair. In her staff email about the layoffs, she said the company had “got ahead of ourselves on hiring” and wrote, “these missteps are on me” REPORTED . That’s rarer than it should be: a founder’s own account, given at the time, of which decisions went wrong. It’s more useful than any case study written afterward.

Mean Business

The cautionary case is older. Al Dunlap’s *Mean Business* (1997) was the playbook of a celebrated turnaround artist. At Sunbeam, the SEC charged, at least $60 million of the company’s reported $189 million in 1997 earnings from continuing operations before taxes came from accounting fraud. In 2002, without admitting or denying the charges, Dunlap agreed to pay a $500,000 civil penalty and to a permanent bar from serving as an officer or director of a public company. He also paid $15 million of his own money to settle a related shareholder class action FILED . Readers who copied his methods were copying a result that, at Sunbeam, the SEC said was partly manufactured.

> **A case study is only as good as the numbers it was built on, and it was written after the ending was known.**

What to take, and what not to

- **Read the filings, not the profiles.** Casper’s losses were in its filings from the start, and Allbirds reported its revenue decline year by year. Where a company is private, look for the founder’s own dated statements, like Glossier’s.
- **Don’t flip the halo.** “Brand-led DTC doesn’t work” is the same mistake as “brand-led DTC wins,” made later. The lesson is about judging from outcomes, not a verdict on the model.
- **Ask Rosenzweig’s question in both directions.** If you’d admired a practice when the brand was growing, decide whether it was a good practice before you learn how the brand is doing now.

Do this

- **Pick the brand your team cites most often** as a model and spend thirty minutes on its filings or its founders' dated statements. Write down one thing the profile left out.

PART THREE · DECIDING BEFORE YOU KNOW · CHAPTER 9

START FROM THE BASE RATE

Before you forecast from your plan, look at what happened to plans like yours. Then move from there, not from the plan.

Daniel Kahneman once helped write a new curriculum. Partway through, he asked the team how long it would take to finish. The estimates ranged from 18 to 30 months. Then he asked the curriculum expert on the team how long similar projects had taken. About 40% had given up, he said, and of the rest, he couldn't think of one finished in less than seven years or more than ten. The project was finished about eight years later PUBLISHED .

The inside view and the outside view

Kahneman and Dan Lovallo told that story in a 1993 paper that named the problem. The inside view forecasts "by focusing on the case at hand": the plan, its obstacles, scenarios of how it will unfold. The outside view "essentially ignores the details of the case at hand" and

looks instead at the statistics of a class of similar cases PUBLISHED . The expert knew the base rate. He still gave an inside-view estimate until someone asked the other question.

Bent Flyvbjerg turned the outside view into a method, reference-class forecasting, and built a database of more than 16,000 big projects. In it, only 8.5% came in on budget and on time, and 0.5% on budget, on time and with the benefits promised PUBLISHED . Your launches aren't bridges. But the pattern of plans beating results is the same, and so is the fix.

> **Your plan is one scenario. The reference class is what usually happens.**

Reference classes for a store

- **New products:** your own last ten launches, their first-90-day units against plan.
- **New channels:** what your last three channel tests cost per order in their first two months against what the plan said.
- **New markets, wholesale accounts, big collaborations:** your own history if you have it; peers' honest numbers if you can get them; public filings for the big versions.

Your own history is the best reference class because it shares your team, product and customers. The ratio of actual to plan is often the most useful single number in it: if your launches have delivered 60% of plan on average, multiply the next plan by 0.6 before you buy inventory.

Adjusting from the base rate

The outside view is where you start, not where you stop. Your specifics matter, to the degree your specifics have predicted results before. The adjustment is the same regression as chapter 5: begin at the reference-class median and move toward your own estimate by the share that your past forecasts have tracked results. If they've tracked well, move most of the way. If they haven't, barely move DERIVED .

RUN YOUR NUMBERS

Adjust a forecast to the reference class

Example numbers. Replace with yours: your forecast for the new launch, and what similar launches actually did. Revenue or units both work.

Your forecast

$ 400000

Similar launches: the middle result

$ 120000

Similar launches: 1 in 5 did worse than

$ 50000

Similar launches: 1 in 5 did better than

$ 250000

How well your past forecasts tracked results, 0 to 1

0.3

$172,205	$53,501 to $554,283	18%	$93,079
adjusted forecast: the middle outcome	likely range: 8 in 10 launches like this land here	chance of reaching your forecast of $400,000	a figure you'd beat 3 times in 4: plan cash and inventory here

> **Your forecast is 2.3 times the adjusted figure, with about 18% chance of reaching it. Keep it as the stretch goal if you like, but buy inventory and set spend for $172,205, and plan cash so you survive $93,079.**

Outcomes like launch revenue are skewed, so the math works on logarithms: it starts at the middle result and moves toward your forecast by the tracking figure. Use 0 if your forecasts have been no better than the base rate, about 0.3 if you don't know, and 0.6 or more only if your records show forecasts tracking results closely. The correlation from chapter 6 is a good estimate.

With the defaults, a team forecasting $400,000 for a launch whose reference class usually does $120,000, and whose forecasts have tracked results only loosely, should plan on about $172,000. Eight in ten launches like it land between about $54,000 and $554,000, and the $400,000 plan has about an 18% chance of happening. That doesn't mean don't launch. It means don't buy inventory for $400,000.

Do this

- **Build your reference class this week:** every launch in the last three years, with its plan and its actual first-90-day result. Compute the median ratio of actual to plan. Put it in the

header of the next launch plan.

PART THREE · DECIDING BEFORE YOU KNOW · CHAPTER 10

KILL CRITERIA

Write down what would make you stop, and when you'll check, before the money is spent and the team is attached.

Every project is easiest to judge before it starts. Nobody's reputation is tied to it yet, no money is sunk, and the team can still say what a failure would look like without it sounding like a verdict on anyone. After launch, all of that changes, and the question "should we keep going?" gets answered by hope.

Decide the exit on the way in

Annie Duke's *Quit* (2022) is a book-length argument for this. Among its tools are quitting contracts, set up in advance, and a warning about escalation of commitment: the pull to keep investing in something because you already have REPORTED . Kill criteria are the operator's version: a written condition, a date and a decision-maker, agreed before launch.

> **The best time to decide when to stop is before you start. The second best is now.**

What a good kill criterion has

01 **A metric the project controls**

Cost per first order from the new channel, not store revenue. Sell-through of the new SKU at full price, not total units.

02 **A threshold built on the base rate**

From the reference class in chapter 9, not from the plan. If similar launches deliver 60% of plan, a kill line at 90% of plan kills good launches.

03 **A minimum sample**

The number of orders or weeks before the criterion counts, so noise can't trigger it. Use the ranges in chapter 2: a threshold set inside the chance range of your sample is a coin flip.

04 **A date**

When it will be checked, on the calendar, with a name next to it.

05 **Three outcomes, not two**

Kill, continue, or scale, each with its own line, so the middle result has a rule too.

A worked example

Say a brand with a $60 contribution per first order launches a new paid channel with $30,000 over eight weeks. Its last three channel tests reached a cost per first order about 40% above the plan in their first two months. The plan says $50. The kill criteria, written before launch:

- **Check:** end of week 6, if the channel has at least 150 first orders; otherwise end of week 8.
- **Kill** if cost per first order is above $85: worse than the reference class by a margin chance can't explain at 150 orders, and well above the contribution.
- **Scale** by half the budget again if it's below $60.
- **Continue at the same budget for four more weeks** between $60 and $85, then decide with no further extension.
- **Decides:** the head of growth. The founder can overrule only in writing, in the decision journal.

Nothing here stops the team from learning. It stops the team from arguing about the threshold in week 6, when everyone already knows which answer they want.

Do this

- **List every live initiative** that costs more than a month of one person's time or a meaningful slice of budget. Write kill criteria for each using the template in Appendix B. For those already running, set the date and threshold now, from the base rate, as if you were starting today.

PART THREE · DECIDING BEFORE YOU KNOW · CHAPTER 11

THE PREMORTEM

Before launch, assume it failed and write down why. Twenty minutes that surface the risks everyone saw and nobody said.

Gary Klein described the method in Harvard Business Review in 2007. "A premortem is the hypothetical opposite of a postmortem." Instead of asking what might go wrong, the leader tells the team the project "has failed spectacularly" and asks why PUBLISHED .

How it runs

01 **Brief the plan**

Everyone hears the final version, including the kill criteria.

02 **Announce the failure**

“It’s six months from now. This launch failed badly.” Make it certain, not possible.

03 **Write alone**

Each person spends a few minutes writing every reason they can think of, “especially the kinds of things they ordinarily wouldn’t mention,” in Klein’s words.

04 **Go around the room**

One reason per person per turn, starting with the project lead, until every reason is on the list.

05 **Change the plan**

The project lead reviews the list and strengthens the plan. Anything that can’t be fixed becomes a kill criterion or a signal to watch.

PUBLISHED Klein, “Performing a Project Premortem,” Harvard Business Review, September 2007.

What the research found, accurately

Klein’s article cites a 1989 study by Deborah Mitchell, Jay Russo and Nancy Pennington, and says it found that imagining an event has already happened “increases the ability to correctly identify reasons for future outcomes by 30%” PUBLISHED . That’s not quite what the study measured. It counted reasons, not correct ones. People told an outcome was certain gave longer explanations, more tied to specific events, than people told it was only possible; a 2010 paper co-written by Klein summarized the effect as about 30% more reasons PUBLISHED . The 1989 paper’s own abstract adds that certainty, not setting the event in the future or the past, drove the difference. The quoted figure is about quantity, not accuracy.

Later tests are small but encouraging. In that 2010 study, by Beth Veinott, Klein and Sterling Wiggins, with 178 students judging a flu-response plan, groups that ran a premortem lowered their confidence in it by 25 points on average, against 12 to 14 points for groups that listed pros and cons or only cons PUBLISHED . More reasons and less overconfidence is

what a launch plan needs. Treat it as a cheap habit with a sound mechanism, not a proven 30% gain.

> **Certainty loosens tongues. “It failed” gets more honest answers than “could it fail?”**

Making it work in a small team

- **Include the people closest to the work:** customer service, the 3PL contact, the designer. They know the failure modes the plan doesn’t.
- **The founder writes too, and speaks last.** Otherwise the list becomes the founder’s list.
- **Keep the list.** It goes in the decision journal, and at the review you’ll see which failures you predicted.

Do this

- **Book twenty minutes** before your next launch, promotion or channel test. Run the five steps, and change at least one thing in the plan because of what comes out. If nothing changes, you ran it too politely.

PART THREE · DECIDING BEFORE YOU KNOW · CHAPTER 12

THE DECISION JOURNAL

Write down what you expected, and why, at the time. It’s the only way to learn from decisions instead of from luck.

A good decision can turn out badly and a bad one can turn out well. If you judge decisions only by outcomes, you'll learn the wrong lessons from both, and you'll learn them confidently.

Process over outcome

Annie Duke, a former professional poker player, built *Thinking in Bets* (2018) on this idea. Her summary: "The quality of our lives is the sum of decision quality plus luck" REPORTED . You control the first term. The second arrives on its own schedule.

	GOOD OUTCOME	BAD OUTCOME
Good decision	Deserved success. Repeat the process.	Bad luck. Keep the process; check the risk was sized right.
Bad decision	Dumb luck. The most dangerous box: it teaches the wrong lesson.	Deserved. Fix the process.

Without a record, everything gets filed by outcome, and the top-right and bottom-left boxes disappear. The dumb-luck box is where overconfident habits come from.

> **Judge the decision on what you knew and how you decided, not on how it turned out.**

What to write

One entry per decision that matters: a launch, a big buy, a hire, a channel, a price change. Written before the result, dated, short.

- **The decision and the options** you considered, including doing nothing.
- **What you expect,** as a number with a range: "first-90-day revenue $150,000 to $250,000, most likely $180,000."

- **How confident you are,** as a percentage.
- **Why:** the two or three reasons that decided it, and the base rate you started from.
- **What would change your mind:** the kill criteria, or the signal that would reopen it.
- **A review date.**

The review

Once a quarter, read the entries whose review dates have passed. For each, put it in a box of the table above. Then look across them. Are your ranges too narrow? If fewer than about eight in ten results fell inside ranges you called 80% likely, they are. Do you overrate one kind of reason? Which calls turned out right for reasons you didn't write down? That's where the learning is, and it's invisible without the page.

Do this

- **Start the journal today** with the template in Appendix B, and write entries for the two biggest decisions currently open. Put the first quarterly review on the calendar.

PART FOUR · RUNNING IT · CHAPTER 13

THE SIGNAL SCORECARD

One page, every week. Each number with its limits and its denominator, the live bets against their kill criteria, and the decisions due for review.

Most weekly scorecards show this week, last week and the change between them. That’s the format most likely to produce a story about noise. The signal scorecard shows each number against its own limits, so the page tells you where to look before anyone starts explaining.

NUMBER	SHOWN AS	READ
Revenue	Process behavior chart, limits from 12 normal weeks; promotion weeks charted separately	Weekly
Orders and average order value	A chart each, so you can see which one moved	Weekly
Site conversion rate	Chart, with sessions beside it	Weekly
Cost per first order	Chart, by channel for the two biggest	Weekly
Revenue per recipient, campaigns and top flows	Chart each; any series under about 100 orders a week marked “read monthly”	Weekly, acted on monthly
Repeat rate, latest full cohort	Chart by cohort month, with the cohort size	Monthly
Live bets	Each one’s metric against its kill line, with the check date	At each check date
Decisions due for review	Journal entries past their review date	Quarterly

The meeting, in four parts

01 **Signals**

Any number that broke a rule. Each gets an owner and one question: what happened that week? Not “what do we change?”

02 **Noise, noted**

Everything else, read out in one sentence: “inside the limits.” No discussion. This is the hardest part of the meeting and the one that saves the most time.

03 **Bets**

Any live bet whose check date has arrived. Apply the kill criteria as written.

04 **Decisions**

Any decision made this week that belongs in the journal. Write the entry before the meeting ends.

> **No explanations for numbers inside their limits. Not even good ones.**

Two rules for the page

First, no rate appears without its denominator. A conversion rate that rose because sessions fell is a different story from one that rose with sessions flat, and the page should show which. Second, every chart carries the dated notes from chapter 3: launches, price changes, flow rewrites, site changes. When a signal appears, the first place to look is the nearest note.

Do this

- **Rebuild next Monday’s scorecard in this format,** even if only three numbers have charts. Run the meeting in the four parts and time it. Keep the timing; it’s the easiest way to show the team what the format saved.

PART FOUR · RUNNING IT · CHAPTER 14

THE FIRST THIRTY DAYS

Limits, then winners, then forecasts and kill criteria, then the premortem and the journal. Four weeks, in that order.

Start with what the team looks at every week, because that's where the most time goes to noise. Then the habits that need the new charts to work. By day thirty, the Monday meeting runs on signals and every live bet has a written exit.

01

Week one: the limits

Chart revenue, orders, conversion rate, revenue per recipient and the latest cohort's repeat rate with 12 normal weeks each (chapter 3). Add denominators to every rate, label small samples, and take open rate out of decisions (chapter 2). Freeze any flow with more than two bad-week edits (chapter 4).

02

Week two: the winners

Run the meeting in the four parts (chapter 13). Put your last three winning campaigns or SKUs through the tool in chapter 5 and resize any budget or reorder built on the headline figure. Run the persistence test on your biggest allocation decision (chapter 6).

03

Week three: forecasts and exits

Build the reference class of your past launches and compute the median ratio of actual to plan (chapter 9). Write kill criteria for every live bet (chapter 10). Answer the six questions for any tactic you're about to copy (chapter 7).

04 **Week four: the premortem and the journal**

Run a premortem on your next launch (chapter 11). Start the decision journal with the two biggest open decisions, and book the first quarterly review (chapter 12).

At day thirty, you won't know yet whether your decisions are better. That takes a quarter of journal entries and a few kill dates. What you'll have is a meeting that spends its time on signals, forecasts that start from what usually happens, and a written record that will show you, in ninety days, where your judgment is good and where it's lucky.

> **Limits before stories. Base rates before plans. Records before hindsight.**

Do this

- **Put the four weeks on the calendar today,** with an owner for each line. Give the owner of week one this guide's Appendix A.

CLOSE

DAY ONE

What whoever runs the weekly numbers needs on the first day.

Whoever owns the scorecard, a new analyst, an operator, or you on the Monday you decide to stop explaining noise, needs six things on day one. Without them, the first month goes on hunting for things they should have been handed.

01 **A year of weekly numbers**

Revenue, orders, sessions, conversion, cost per first order, and email and SMS revenue per recipient, by week, in one sheet.

02 **The promotion and change calendar**

Every sale, launch, price change, flow rewrite and site change of the last year, with dates. Gaps are findings.

03 **Campaign history**

Every campaign of the last year with its audience size, orders and revenue, for the spread in chapter 5.

04 **Launch history**

Every product, channel and market launch of the last three years, with its plan and its actual result.

05 **The list of live bets**

Everything now running that costs real money or time, with its owner.

06 **Authority to say “noise”**

The founder’s agreement that a number inside its limits gets no explanation in the meeting.

Do this

- **Ask for all six in one message** today. Whatever comes back slowest is where the business knows least about itself.

CLOSE

THE SHELF

The books worth reading next, and what to take from each.

Understanding Variation

Donald J. Wheeler, 1993

A short book that teaches the process behavior chart to managers, with no statistics background assumed. The source for part one.

The Success Equation

Michael J. Mauboussin, 2012

How to separate skill from luck in any field, and what to do differently depending on which one dominates.

The Halo Effect

Phil Rosenzweig, 2007

Why most business bestsellers are built on outcomes dressed as causes. Read it before your next case study.

Thinking, Fast and Slow

Daniel Kahneman, 2011

The chapters on regression to the mean and the outside view are the best short treatment of both.

How Big Things Get Done

Bent Flyvbjerg and Dan Gardner, 2023

Reference-class forecasting from the person who built it into a method, with the evidence from thousands of projects.

Quit

Annie Duke, 2022

On deciding in advance when to walk away, and why it's so hard to do once you're in.

Hard Facts, Dangerous Half-Truths and Total Nonsense

Jeffrey Pfeffer and Robert I. Sutton, 2006

The case for deciding from the best evidence about what works, the way medicine learned to, instead of from management fads and half-truths.

And the research: Galton (1886) on regression; Kahneman and Lovallo (1993) on the inside and outside view; Mitchell, Russo and Pennington (1989) and Veinott, Klein and Wiggins (2010) on premortems; Denrell (2003) on survivors; Golder and Tellis (1993) on pioneers; Barber and Odean (2000, 2001) on overtrading. Full references are in Appendix C.

CLOSE

ABOUT THE AUTHOR

Andrew Lauchner is a growth and retention operator for consumer brands and the author of sixteen free books on the method. He fixes the second-purchase problem through his practice, Growth Legend, as one engagement: The Second-Purchase Sprint. He is the author of The Second Order, on turning first-time buyers into second-time buyers, and The Whole Machine, on the fundamentals of DTC growth, along with a series of field guides for DTC operators at andrewlauchner.com.

As Senior Director of Growth and Retention Marketing at Gallery Furniture, he rebuilt the customer journey and the sales playbooks together. He has worked on growth and retention at Binance and 3Commas, and has been Head of Growth and Retention at Greatness Wins and at Nexus Agriscience.

WHAT COLLEAGUES SAY

“Andrew led retention, lifecycle, and email/SMS, but what separates him from most in this space is how deeply he understands the role retention plays in the overall growth engine.”

Akram Khan, Head of Marketing at Gallery Furniture, senior to Andrew but didn’t manage Andrew directly

Andrew answers every note from operators working on this. Write to andrew@growthlegend.com or message him on LinkedIn.

APPENDIX A

FOR YOUR ANALYST

The formulas behind the three tools, and the queries that feed them.

The formulas

FOR	FORMULA	NOTES
Natural process limits	$\bar{X} \pm 2.66 \times m\bar{R}$	$\bar{X}$: average of the baseline weeks. $m\bar{R}$: average of $\lvert X_t - X_{t-1} \rvert$. 2.66 = 3 / 1.128.
Moving range limit	$3.268 \times m\bar{R}$	Only rule for the moving ranges: a point above it.
Signal rules	1 point outside; 3 of 4 beyond $\bar{X} \pm 1.33\, m\bar{R}$; 8 in a row on one side	1.33 $m\bar{R}$ is halfway from the central line to a limit.
Chance-alone range	$1.96 \times \sqrt{p(1 - p) / n}$; counts: $1.96 \times \sqrt{n}$	95% range from sampling alone.
Share of a lead that holds	$w = (s^2 - \bar{p}(1 - \bar{p})/m) / (s^2 - \bar{p}(1 - \bar{p})/m + \bar{p}(1 - \bar{p})/n)$	s: spread of past rates; m: typical past audience; n: the winner's audience. Floor the numerator at 0. Expected = $\bar{p}$ + w(observed − $\bar{p}$).
Reference-class adjustment	$\mu = \ln(\text{median}) + r(\ln(\text{forecast}) - \ln(\text{median}))$	σ = (ln P80 − ln P20) / 1.683; residual $\sigma\sqrt{(1 - r^2)}$. Range: $\exp(\mu \pm 1.2816 \times \text{residual})$. Plan figure: $\exp(\mu - 0.6745 \times \text{residual})$.

Weekly series, with limits

```
-- weekly revenue and orders, excluding promotion weeks from the baseline
-- orders: order_id, created_at, total_price, cancelled_at
-- promo_weeks: week_start (one row per promotion week)
```

```
WITH weekly AS (
  SELECT DATE_TRUNC('week', created_at)::date AS week_start,
         COUNT(*)                             AS orders,
         SUM(total_price)                     AS revenue
  FROM orders
  WHERE cancelled_at IS NULL
  GROUP BY 1
),
base AS (
  SELECT w.*, ABS(revenue - LAG(revenue) OVER (ORDER BY week_start)) AS mr
  FROM weekly w
  WHERE week_start NOT IN (SELECT week_start FROM promo_weeks)
    AND week_start < DATE_TRUNC('week', CURRENT_DATE)
  ORDER BY week_start DESC
  LIMIT 12
)
SELECT AVG(revenue)                     AS central_line,
       AVG(revenue) - 2.66 * AVG(mr)    AS lower_limit,
       AVG(revenue) + 2.66 * AVG(mr)    AS upper_limit,
       3.268 * AVG(mr)                  AS moving_range_limit
FROM base;
```

The moving ranges here are computed before the twelve weeks are chosen, so the oldest week's range reaches back one normal week; that's fine. If a promotion week sits between two normal weeks, the range spans it, which is also fine. Swap `revenue` for `orders` or a rate to chart the others. Refunds belong in a separate series from the `refunds` table, not netted into revenue, so a returns problem shows on its own chart.

Campaign spread, for the estimator

```
-- last 30 email campaigns: order rate per recipient, its spread, typical audience
-- campaign_sends: campaign_id, channel, sent_at, recipients
-- campaign_orders: campaign_id, order_id (orders attributed to the campaign)
WITH c AS (
  SELECT s.campaign_id, s.recipients,
         COUNT(o.order_id)::numeric / NULLIF(s.recipients, 0) AS rate
  FROM campaign_sends s
  LEFT JOIN campaign_orders o USING (campaign_id)
  WHERE s.channel = 'email'
  GROUP BY s.campaign_id, s.recipients, s.sent_at
  ORDER BY s.sent_at DESC
  LIMIT 30
)
SELECT AVG(rate) * 100          AS avg_rate_pct,
```

```
       STDDEV_SAMP(rate) * 100  AS spread_pts,
       AVG(recipients)          AS typical_audience
FROM c;
```

Use the same attribution window for every campaign. Mixing sends to the whole list with sends to small segments inflates the spread; run it separately for each.

Launch reference class

```
-- first-90-day revenue against plan, for every launch
-- launches: launch_id, product_id, launched_on, plan_revenue_90d
-- order_lines: order_id, product_id, price, quantity; orders: order_id, created_at
WITH actual AS (
  SELECT l.launch_id, l.plan_revenue_90d,
         SUM(ol.price * ol.quantity) AS revenue_90d
  FROM launches l
  JOIN order_lines ol ON ol.product_id = l.product_id
  JOIN orders o       ON o.order_id = ol.order_id
                     AND o.created_at >= l.launched_on
                     AND o.created_at <  l.launched_on + INTERVAL '90 days'
  WHERE l.launched_on < CURRENT_DATE - INTERVAL '90 days'
  GROUP BY 1, 2
)
SELECT PERCENTILE_CONT(0.2) WITHIN GROUP (ORDER BY revenue_90d) AS p20,
       PERCENTILE_CONT(0.5) WITHIN GROUP (ORDER BY revenue_90d) AS median,
       PERCENTILE_CONT(0.8) WITHIN GROUP (ORDER BY revenue_90d) AS p80,
       PERCENTILE_CONT(0.5) WITHIN GROUP
         (ORDER BY revenue_90d / NULLIF(plan_revenue_90d, 0))    AS median_actual_to_plan
       CORR(LN(plan_revenue_90d), LN(revenue_90d))              AS forecast_tracking
FROM actual
WHERE revenue_90d > 0 AND plan_revenue_90d > 0;
```

`forecast_tracking` is the tracking figure for the tool in chapter 9. With fewer than about ten launches it's rough; round it down.

Rank persistence

```
-- does last quarter's ranking by cost per first order hold this quarter?
-- channel_quarters: channel, quarter, spend, first_orders
WITH r AS (
  SELECT channel, quarter,
```

```
         RANK() OVER (PARTITION BY quarter
                      ORDER BY spend / NULLIF(first_orders, 0)) AS rnk
  FROM channel_quarters
  WHERE quarter IN ('2026-Q1', '2026-Q2') AND first_orders >= 30
)
SELECT CORR(a.rnk, b.rnk) AS rank_persistence
FROM r a
JOIN r b ON a.channel = b.channel
WHERE a.quarter = '2026-Q1' AND b.quarter = '2026-Q2';
```

The same query works for creators, ad sets or campaign types; change the table. The first-orders floor keeps tiny channels, whose rank is almost pure chance, out of the comparison.

APPENDIX B

TEMPLATES

Five one-page forms. Copy them into whatever your team already uses.

Kill criteria

```
BET               what we're launching, and its budget:
OWNER             who runs it:
DECIDES           who applies the criteria (and who can overrule, in writing):
METRIC            one number the bet controls:
BASE RATE         what similar bets did (reference class, chapter 9):
MINIMUM SAMPLE    orders or weeks before the criteria count:
CHECK DATE        on the calendar:
KILL IF           metric worse than:
```

```
SCALE IF          metric better than:
CONTINUE IF       in between, for how long, and then decide with no extension:
```

The premortem

```
20 MINUTES, BEFORE LAUNCH
 2 min   Brief the final plan and its kill criteria.
 1 min   "It's [date, six months out]. This failed badly."
 6 min   Everyone writes every reason, alone. Include the ones
         you wouldn't normally say.
 8 min   Round the room, one reason per turn, until the lists
         are empty. The founder goes last.
 3 min   Owner picks the reasons that change the plan today,
         and the ones that become signals to watch.
AFTER    List goes in the decision journal entry for this launch.
```

Decision journal entry

```
DATE              today:
DECISION          what we chose:
OPTIONS           what else we considered, including doing nothing:
BASE RATE         what usually happens in cases like this:
EXPECT            a number and a range (low / most likely / high):
CONFIDENCE        % that the result lands in the range:
WHY               the two or three reasons that decided it:
CHANGE MY MIND    the signal or kill criterion that would reopen it:
REVIEW DATE       when we'll read this again:
-- at review --
RESULT            what happened:
BOX               good or bad decision / good or bad outcome:
LESSON            what we'd do differently, about the process:
```

The Monday agenda

```
1. SIGNALS      numbers that broke a rule. Owner and one question:
                "what happened that week?"
2. NOISE        everything else, read out as "inside the limits."
                No discussion.
3. BETS         live bets at their check date. Apply the criteria.
```

```
4. DECISIONS    anything decided this week goes in the journal
                before the meeting ends.
```

Before we copy it

```
TACTIC              what, and whose:
1. CAUSE            did it cause their success, or come with it?
2. FAILURES         who did the same and failed? (name them, or say
                    you didn't look)
3. HALO             would we rate it the same if they were struggling?
4. CONDITIONS       what made it work there that we don't share?
5. SMALL TEST       cheapest way to find out here:
6. DOWNSIDE         what it costs if we're wrong:
```

APPENDIX C

SOURCES

Every external source, by chapter. Web sources were read in September 2026.

Noise and control charts (chapters 2 to 4)

- Walter A. Shewhart. Economic Control of Quality of Manufactured Product. 1931. ASQ, Walter A. Shewhart, on the memorandum of May 16, 1924. On chance, assignable, common and special causes: overview.
- Donald J. Wheeler. Understanding Variation: The Key to Managing Chaos. SPC Press, 1993. "What Makes the XmR Chart Work?" Quality Digest Daily, December 3, 2012.

- On the three detection rules and baseline length: ProSocia, “Process behavior charts are the key to understanding your business”; Deming Alliance, “Process Behaviour Charts: An Introduction”. Mark Graban, Measures of Success.
- The W. Edwards Deming Institute. On process tampering and the funnel experiment.
- Apple. Mail Privacy Protection and privacy.
- Brad M. Barber and Terrance Odean. “Trading Is Hazardous to Your Wealth: The Common Stock Investment Performance of Individual Investors.” Journal of Finance 55(2), 2000, 773–806.
- Brad M. Barber and Terrance Odean. “Boys Will Be Boys: Gender, Overconfidence, and Common Stock Investment.” Quarterly Journal of Economics 116(1), 2001, 261–292.

Regression, skill and luck (chapters 5, 6)

- Francis Galton. “Regression Towards Mediocrity in Hereditary Stature.” Journal of the Anthropological Institute 15, 1886, 246–263. Data: HistData.
- Daniel Kahneman. Autobiography, Nobel Prize, 2002. Thinking, Fast and Slow, 2011, chapter 17; quoted in Farnam Street.
- Michael J. Mauboussin. The Success Equation: Untangling Skill and Luck in Business, Sports, and Investing. 2012; quoted in Farnam Street. “The True Measures of Success.” Harvard Business Review, October 2012. Interview, Farnam Street.

Playbooks and case studies (chapters 7, 8)

- Phil Rosenzweig. The Halo Effect. 2007. “The halo effect, and other managerial delusions.” McKinsey Quarterly, February 2007.
- Jerker Denrell. “Vicarious Learning, Undersampling of Failure, and the Myths of Management.” Organization Science 14(3), 2003, 227–243. “Selection Bias and the Perils of Benchmarking.” Harvard Business Review, April 2005.
- Peter N. Golder and Gerard J. Tellis. “Pioneer Advantage: Marketing Logic or Marketing Legend?” Journal of Marketing Research 30(2), 1993, 158–170.
- Jeffrey Pfeffer and Robert I. Sutton. “Evidence-Based Management.” Harvard Business Review, January 2006. Hard Facts, Dangerous Half-Truths and Total Nonsense. 2006.
- Casper Sleep Inc. IPO price from its 10-K for 2020 and 10-Qs (SEC XBRL data); revenue and net loss from its 10-K for 2020, filed February 26, 2021; “Durational Capital Management Completes Acquisition of Casper,” 8-K exhibit 99.1, January 25, 2022.
- Allbirds, Inc. “Allbirds Announces Pricing of Initial Public Offering,” November 2, 2021. Net revenue from 10-Ks for 2021 to 2025 (SEC XBRL data). Smartbird, Inc. (formerly Allbirds, Inc.), 10-Q for the quarter ended June 30, 2026.
- Madeline Halpert. “Emily Weiss Steps Down As Glossier CEO Months After Company Layoffs.” Forbes, May 24, 2022.
- Albert J. Dunlap. Mean Business: How I Save Bad Companies and Make Good Companies Great. 1997. SEC, Litigation Release No. 17710, SEC v. Albert J. Dunlap et al., September 2002.

Deciding before you know (chapters 9 to 12)

- Daniel Kahneman and Dan Lovallo. “Timid Choices and Bold Forecasts: A Cognitive Perspective on Risk Taking.” Management Science 39(1), 1993, 17–31.
- Bent Flyvbjerg and Dan Gardner. How Big Things Get Done. 2023. Figures as summarized in The Independent Review, Fall 2023.
- Annie Duke. Quit: The Power of Knowing When to Walk Away. 2022. Thinking in Bets: Making Smarter Decisions When You Don’t Have All the Facts. 2018. Books and Thinking in Bets pages.
- Gary Klein. “Performing a Project Premortem.” Harvard Business Review, September 2007.
- Deborah J. Mitchell, Jay Russo and Nancy Pennington. “Back to the future: Temporal perspective in the explanation of events.” Journal of Behavioral Decision Making 2(1), 1989, 25–38.
- Beth Veinott, Gary A. Klein and Sterling Wiggins. “Evaluating the Effectiveness of the PreMortem Technique on Plan Confidence.” Proceedings of the 7th ISCRAM Conference, 2010. Jason Collins, “The premortem,” course notes, on the 30% figure.

Filed figures come from securities filings and SEC releases, as of the dates given. Published figures come from peer-reviewed research and from books and papers by the researchers themselves. Reported figures come from named outlets, companies and practitioners. Derived figures are my arithmetic with a standard formula, inputs shown. Examples that open with Say or Picture use made-up round numbers. The investing research is cited for what it shows about decisions, not as investment guidance. None of this is legal or financial advice.

Privacy