20 *Why Good Analyses Go Wrong: Behavioral Pitfalls and Forecasting Discipline*

After this chapter you will be able to:

1. state the average constant-price cost overruns the record reports for rail, for bridges and tunnels, and for roads, and explain why these errors are bias and not noise
2. distinguish optimism bias from strategic misrepresentation, and match each one to its cure
3. recognize anchoring in a review that begins from the advocate's own estimate, and state what an independent review does instead
4. decide whether to finish or abandon a part-built project using only future cash flows
5. diagnose confirmation bias in a one-sided sensitivity table, and complete the table in both directions
6. compute budgets at the 50% and 80% confidence levels by applying the reference class uplifts

Every method in this book is exact. Present worth, annual equivalence, rate of return, the replacement rules: given the cash flows, the arithmetic never lies. The cash flows are the problem. They are estimates, made by people, and people estimate with systematic, measurable, directional error. Chapter 1 flagged this in a margin note and promised a full treatment; this chapter is that treatment. It is built on evidence, not exhortation: the largest statistical studies of forecast accuracy in capital projects, summarized in Flyvbjerg (2006), whose numbers Chapter 1's case study introduced. The chapter first reads the planned-versus-actual record

and the two explanations that survive testing, optimism bias and strategic misrepresentation. It then names the individual habits that corrupt analyses from inside, anchoring, sunk-cost escalation, confirmation bias in sensitivity work, and presents the correction that has moved into government practice: reference class forecasting, the outside view. It closes with the working disciplines, pre-mortems, independent estimate reviews, and governance that separates the estimator from the advocate, and with the point of it all: the audit mindset this book has been training since its first ''assessing outside recommendations'' problem.

Why careful estimates still miss.

Q1. A metro rail project is budgeted at 200 million SAR. Across the largest published samples, rail projects overran their constant-price budgets by 44.7% on average. If this project performs at the historical average, what will it cost, and why is ''our engineers are better than average'' not an automatic defense?

Q2. A half-built plant has consumed 300 million SAR. Finishing it will take another 250 million, and the finished plant's future benefits are worth 280 million in present worth. Should it be finished, and what role does the 300 million play?

Q3. What is the difference between the inside view and the outside view of a project forecast?

Answers.

Q1. $200 \times 1.447 \approx 289$ million SAR. The defense fails because the studied forecasts were also made by competent engineers who also believed they were better than average; the record shows the error is a stable property of how estimates are produced, not of who produces them. Only demonstrated, documented estimating accuracy justifies a smaller allowance.

Q2. Finish: the future comparison is $280 > 250$, a gain of 30 million. The 300 million plays no role at all; it is sunk, identical in both futures (chapter 16). It is also no *argument* for finishing: had completion cost 300 million instead, the right answer would be to stop, however painful the write-off.

Q3. The inside view builds the forecast from the project's own components: its design, its plan, its imagined future. The outside view ignores the project's specifics and asks how projects *like it* actually turned out, then places this one in that distribution. The

record shows the outside view is more accurate, because it cannot be reached by the biases that shape the inside one.

20.1 *The Planned-versus-Actual Record*

Begin with what actually happens. Flyvbjerg (2006) summarizes the first large statistical studies of cost and demand forecasts in transportation infrastructure, hundreds of projects spanning seven decades, with inaccuracy measured against the forecast made at the decision to build, in constant prices, so inflation is already excluded.

Reference. Flyvbjerg (2006), *Project Management Journal*, 37(3), 5–15; open access at arXiv:1302.3642. All figures in this chapter are from that paper unless stated.

Forecast inaccuracy. Measured as (actual/forecast − 1) × 100%, with the forecast taken at the decision to build, in constant prices. Zero means a perfect forecast.

project type	average cost overrun	standard deviation
rail	+44.7%	38.4
bridges and tunnels	+33.8%	62.4
roads	+20.4%	29.9

Table 20.1. Inaccuracy of construction-cost forecasts in constant prices (Flyvbjerg, 2006, Table 1). All three averages differ from zero with very high statistical significance.

Demand forecasts err too, and not symmetrically: actual rail passenger traffic averaged 51.4% *below* forecast (standard deviation 28.1), with 84% of rail projects wrong by more than ±20% and 72% wrong by more than ±40%; road traffic averaged 9.5% *above* forecast (standard deviation 44.3), with half of road forecasts wrong by more than ±20%. Costs run over while rail demand runs under, the two errors compounding in the direction that flatters the proposal. And over the 70 years for which cost data exist, accuracy has not improved: better models, better data, and better computers have not moved the averages.

Insight. The record shows *bias*, not noise. If forecasts were merely uncertain, their errors would center near zero and shrink as methods improved; instead the distributions are non-normal, the means are far from zero, always in the favorable direction, and seventy years of technical progress have not touched them. That is the signature of a systematic cause, and it is why the correction in this chapter is statistical and institutional rather than technical: a better model cannot fix an estimate whose problem is the estimator's incentives and psychology.

Figure 20.1 puts the three averages of Table 20.1 on one axis; look at where the means sit relative to the zero line, not at how wide the spreads around them are.

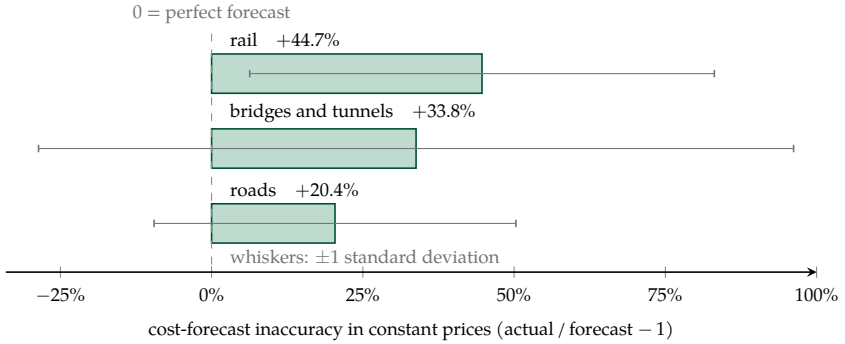

Figure 20.1. Average constant-price cost overruns for the three project types of Table 20.1, with whiskers at one standard deviation. Every mean lies well to the right of the zero line instead of straddling it, which is the signature of bias rather than noise.

Example 20.1 — What the record does to a present worth

Najd Copper evaluates a rail spur from the mine to the port: capital cost 500 million SAR, net benefits 90 million SAR per year for 20 years, MARR 10%, $(P/A,\ 10\%,\ 20) = 8.5136$. (a) Screen the project on the forecast. (b) Re-screen it at the historical averages for rail: cost +44.7%, demand-driven benefits −51.4%. (c) Compare the benefit–cost ratios.

Solution.

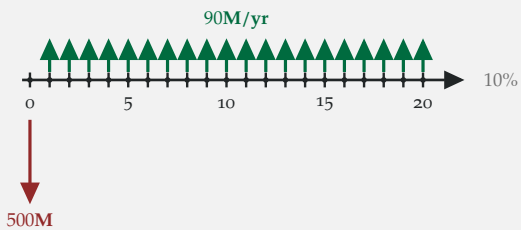

Figure 20.2. The rail spur as forecast: 500 million SAR of capital against twenty years of 90-million benefits. The record's question is whether either number deserves belief.

Step (a): at the forecast.

$$\text{PW} = -500 + 90\,(8.5136) = -500 + 766.22 = \boxed{+266.2 \text{ million SAR}} \Rightarrow \text{accept.}$$

Step (b): at the record's averages.

Cost $500 \times 1.447 = 723.5$; benefits $90 \times (1 - 0.514) = 90 \times 0.486 = 43.74$ per year:

$$\begin{aligned} \text{PW} &= -723.5 + 43.74\,(8.5136) = -723.5 + 372.38 \\ &= \boxed{-351.1 \text{ million SAR}} \;\Rightarrow\; \text{reject.} \end{aligned}$$

Step (c): the ratios.

Forecast: $766.22/500 = 1.53$. At the averages: $372.38/723.5 = 0.51$. The project presented as returning 1.53 units of discounted benefit per unit of cost delivers, at the historical averages, about 0.51, because the cost error inflates the denominator while the demand error deflates the numerator. A 617-million swing in present worth separates the proposal from the record, and nothing in the proposal's own pages reveals it.

20.2 *Optimism Bias and Strategic Misrepresentation*

Errors of this size, persistence, and direction need explaining. Flyvbjerg and colleagues tested three families of explanation against the data. *Technical* explanations, bad data, immature models, fail: they predict errors centered near zero that shrink over time, and the record shows neither. Two explanations survive.

Optimism bias operates even with honest, disinterested estimators: it is the planning fallacy of Kahneman and Tversky, the cognitive tilt that makes one's own project feel like the exception. *Strategic misrepresentation* operates where money and approval are competed for: proposals are shaded because the shaded proposal wins. The two produce the same statistical signature, flattering forecasts, but they respond to different cures, which is why the diagnosis matters. Where political and organizational pressure is low, optimism dominates and better methods are welcomed; where pressure is high, misrepresentation dominates and better methods are resisted, because the inaccuracy is the point. Both usually coexist on large projects.

Optimism bias. The documented human tendency to judge one's own plans more favorably than experience warrants: costs and times underestimated, benefits overestimated. It is self-deception, made in good faith.

Strategic misrepresentation. The deliberate shading of estimates, costs low, benefits high, to win approval or funding in competition with other proposals. It is deception, made knowingly, under organizational pressure.

Insight. Keep the two apart when assigning remedies. Optimism bias yields to method: outside-view data, checklists, reviews, the honest estimator is grateful for them. Strategic misrepresentation yields only to incentives: independent review the promoter cannot choose, budgets that bind the forecaster to the forecast, and consequences for repeated flattery. Handing a methods fix to an incentives problem produces compliance theater: the biased number, now beautifully documented.

Anchoring. The tendency of any stated number to pull subsequent estimates toward itself. Adjustments away from an anchor are systematically insufficient, even when the anchor is known to be arbitrary or interested.

20.3 *Anchoring*

The first number spoken in a meeting does not just inform the discussion; it captures it. Estimates made after an anchor is on the table move toward the anchor, and adjustments away from it stop too soon, a result demonstrated across decades of judgment research and visible in every budget review. For capital projects the anchor is almost always the advocate's own estimate, which is precisely the number optimism and strategy have already shaped. A review that starts from the proposal's figure and adjusts is therefore not independent, whatever its letterhead says; it is the advocate's number wearing a reviewer's signature.

Example 20.2 — A review that never left the anchor

The rail spur of Example 20.1 is submitted at 500 million SAR. A review committee, wanting to be prudent, adds 10% and approves a budget of 550 million. (a) State what anchored the committee. (b) Compare its budget with the reference-class requirement for rail at the 50% and 80% confidence levels (uplifts of 40% and 57%, Section 20.6). (c) Quantify the gap.

Solution.

Step (a): the anchor.

The committee's starting point was the advocate's 500, and its ''prudence'' was an adjustment *from that anchor*, sized by feel. Nothing in the 10% came from data; the record says rail overruns average 44.7%, so an average project blows through the committee's margin four times over.

Step (b): the outside view.

From the uplift table:

$$500 \times 1.40 = 700 \text{ (50\% confidence)},$$
$$500 \times 1.57 = \boxed{785 \text{ million SAR}} \text{ (80\% confidence)}.$$

Step (c): the gap.

The anchored budget of 550 sits 235 million below the 80%-confidence figure, and 150 million below even the coin-flip figure. The committee believed it had been conservative; it had been anchored. The repair is procedural, not attitudinal: the review must *produce its number before reading the advocate's*, from the reference class, and only then reconcile.

20.4 *Sunk-Cost Escalation*

Escalation of commitment. Continuing to fund a failing course of action because of what has already been spent. The sunk amount is irrelevant to the forward decision; using it as a reason is throwing good money after bad.

chapter 16 established the rule: money already spent is identical in every future and can never separate alternatives. The behavioral finding is that people violate this rule predictably and in a known direction, spending *more* readily on a troubled project the more it has already consumed. The psychology is loss aversion dressed as perseverance: stopping converts the sunk amount into an admitted loss, and admitted losses hurt more than escalating commitments seem to. On capital projects the escalation is usually spoken in one of two sentences: ''we cannot waste what we have spent,'' or ''we are almost there.'' Neither sentence contains a cash flow.

Example 20.3 — Finish the filter plant?

Najd Copper's tailings filter plant was approved at 150 million SAR. With 120 million spent, a revised estimate says completion needs 95 million more, and the market for filtered tailings has weakened: the completed plant's future net benefits are now worth 130 million SAR in present worth. Abandoning the site would recover 12 million in salvage. (a) Decide forward: finish or abandon. (b) State what role the 120 million plays. (c) Find the completion cost at which the verdict flips. (d) Evaluate the project as a whole, and reconcile the two verdicts.

Solution.

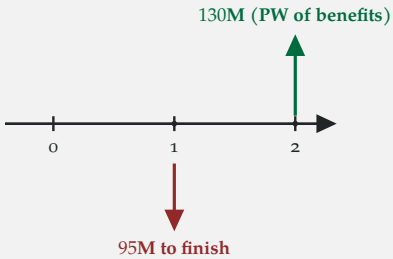

Figure 20.3. The forward decision, drawn without the sunk 120 million: pay 95 to finish, receive future benefits worth 130, or take 12 in salvage. Only these flows exist in the future.

Step (a): forward comparison.

Finish: $-95 + 130 = +35$ million. Abandon: $+12$ million. Since $35 > 12$, finish the plant, by a margin of 23 million.

Step (b): the 120 million.

None. It is sunk, the same in both futures. If anyone in the room says "finish, because 120 million must not be wasted," they have reached the right conclusion by an invalid argument, and the invalid argument will not stay right: it would also say "finish" if completion cost 200 million.

Step (c): the flip point.

Finish while completion cost $< 130 - 12 =$ 118 million SAR. The re-estimate should be monitored against that threshold, not against the original budget.

Step (d): the whole project.

Total outlay $120 + 95 = 215$ million for benefits worth 130: the venture destroyed about 85 million of value, and the original 150-million approval was, in hindsight, built on estimates that failed. Both statements are true at once: *finishing is right* and *starting was wrong*. The first is a decision; the second is an input to Section 20.7's estimate reviews, so the next approval is better. Escalation thrives when firms refuse to hold both thoughts.

Insight. The sunk-cost audit question is always the same: *does the argument for con-*

tinuing contain any future cash flow? ''We've spent too much to stop'' contains none. ''Completion costs 95 against benefits of 130'' contains two. The first family of arguments is noise regardless of conclusion; the second decides.

20.5 *Confirmation Bias in Sensitivity Analysis*

chapter 12 made sensitivity analysis the first defense against uncertain estimates. The behavioral warning is that the defense can be pointed the wrong way. An analyst who wants the project (or whose sponsor does) chooses which inputs to vary and in which directions, and confirmation bias makes those choices flattering: recovery up, price up, capex favorable. The resulting table is arithmetically impeccable and evidentially empty, because a project that is only tested where it cannot fail has not been tested. The tell is asymmetry: swings in one direction, ranges that exclude the historical downside, and the word ''conservative'' attached to inputs that are anchors (Section 20.3) rather than measurements.

In practice. The trigger for the re-estimate should be condition data, not the calendar: measured productivity against plan, quantities installed against design, trade contractor claims. Escalation survives longest on projects where nobody re-measures, because without a fresh forward estimate the sunk narrative is the only number in the room.

Confirmation bias. The tendency to seek, weight, and present evidence that supports a favored conclusion. In project analysis it appears as sensitivity cases chosen from the favorable directions only.

In practice. The adverse directions are engineering knowledge. The failure history of similar ore bodies, the commissioning curves of past plants, the change-order log, these say which way reality usually moves. A sensitivity plan drafted by the operations and maintenance staff will contain the downside cases the project office forgot.

Example 20.4 — A one-sided table, made honest

A memo for Najd Copper's heap-leach add-on (capital 40 million SAR, net benefits 8.2 million per year for 8 years, MARR 10%, $(P/A,\ 10\%,\ 8) = 5.3349$) reports a base present worth of $+3.75$ million and this sensitivity table: recovery $+2$ points gives $+8.12$; price $+10\%$ gives $+8.12$; capital -5% gives $+5.75$. It concludes: ''the project remains attractive in every case examined.'' (a) Verify the base case. (b) Identify the bias. (c) Complete the table symmetrically and restate the conclusion.

Solution.

Step (a): base case.

$$\text{PW} = -40 + 8.2\,(5.3349) = -40 + 43.75 = \boxed{+3.75 \text{ million SAR}}.$$

Step (b): the bias.

Every tested case moves an input in its *favorable* direction: benefits up

(+2 points of recovery or +10% of price each add $0.82 \times 5.3349 = 4.37$), capital down (−5% adds 2.00). The memo's arithmetic is correct; its selection is one-sided, so the sentence ''attractive in every case examined'' describes the examiner, not the project.

Step (c): the symmetric table.

Add the mirror cases:

input swing	PW up (M SAR)	PW down (M SAR)
recovery ±2 points	+8.12	−0.63
price ±10%	+8.12	−0.63
capital ∓10%	+7.75	−0.25

Table 20.2. The heap-leach sensitivity table completed in both directions (capital shown at the symmetric ±10%). Each of the three downside cases, absent from the memo, takes the present worth negative.

Each downside case flips the sign, and a joint mild downside (price −10% with capital +10%) reaches $3.75 - 4.37 - 4.00 = -4.62$ million. Honest restatement: *the project is a knife-edge accept whose verdict reverses under one-input downside swings of ordinary size*; it needs the probabilistic tools of chapter 12 and chapter 14, or better estimates, before approval. Please note the general rule, from Section 12.3: a sensitivity table is complete only when every varied input swings both ways.

Inside vs. outside view. The inside view forecasts from the project's own components and plan. The outside view forecasts from the measured outcomes of a class of comparable completed projects, ignoring the project's specifics.

Reference class forecasting. The three-step outside-view procedure: (1) choose a reference class of comparable past projects; (2) establish the statistical distribution of their outcomes; (3) place the current project in that distribution.

Uplift. The percentage added to a submitted capital estimate so the resulting budget holds at a chosen confidence level, read from the reference class's overrun distribution. Higher confidence demands a higher uplift.

20.6 *Reference Class Forecasting: The Outside View*

The correction that follows from the diagnosis is not ''estimate more carefully.'' Careful estimating is inside-view work, and the inside view is where the bias lives. The correction is to *bypass* the inside view: base the forecast on how projects like this one actually turned out. Kahneman and Tversky's prescription, developed from the decision research that won the 2002 Nobel prize in economics, became practical policy when the UK Department for Transport commissioned empirically based *uplifts*: percentages added to a submitted budget so that the funded amount holds at a stated confidence level, derived from the measured overrun distributions of hundreds of completed projects.

category	50% percentile	80% percentile
roads	15%	32%
rail	40%	57%
fixed links (bridges, tunnels)	23%	55%
building projects	4–51%*	
IT projects	10–200%*	

Table 20.3. Applicable optimism-bias uplifts, constant prices (Flyvbjerg, 2006, Table 4). *Ranges from a prior UK study without full distributions. The 50% level suits portfolio investors whose overruns and underruns offset; the 80% level is what the UK requires for stand-alone funding decisions.

The percentile is a risk choice, not a technical one. A funder with a large portfolio can budget at the 50% level, letting overruns and underruns offset across projects. A stand-alone project with no access to further funds budgets at the 80% level or above, paying a larger reserve for a smaller chance of returning to ask for more. And the uplift may be lowered only on *evidence* that this estimator's track record beats the class, the one defense the method accepts.

In practice. The published classes are transport projects in one country. Your plant's truest reference class is its own history, and building that distribution is engineering work: gathering the records, normalizing them, including the failures honestly. A firm that has done it owns an uplift table no consultant can sell its competitors.

Example 20.5 — Uplifting a Najd Copper capital program

Najd Copper's concentrator expansion includes three infrastructure estimates, submitted at the decision to build: an 18-km mine access road upgrade at 80 million SAR; a rail spur to the port at 150 million SAR; and a plant control-system (IT) project at 25 million SAR. Using Table 20.3, set budgets at the 50% and 80% confidence levels, and state what the IT row's range means for the third estimate.

Solution.

Step 1: the road (roads class: 15% and 32%).

$$80 \times 1.15 = 92, \qquad 80 \times 1.32 = \boxed{105.6 \text{ million SAR}}.$$

Step 2: the rail spur (rail class: 40% and 57%).

$$150 \times 1.40 = 210, \qquad 150 \times 1.57 = \boxed{235.5 \text{ million SAR}}.$$

Step 3: the control system (IT class: 10–200%, no distribution).

$$25 \times 1.10 = 27.5 \quad \text{up to} \quad 25 \times 3.00 = \boxed{75 \text{ million SAR}}.$$

The width is itself the information: the reference class for IT projects is so dispersed that no single uplift is defensible, which argues for staging the project (chapter 13) rather than funding it in one decision.

Step 4: read the totals.

The two civil items submitted at $80 + 150 = 230$ million need $92 + 210 = 302$ million to be even-money and $105.6 + 235.5 = 341.1$ million to be 80%-safe, an uplift of 48% on the combined submission. Two cautions attach. First, the percentages come from UK transport classes, the nearest published relatives, not from Saudi mining infrastructure; the firm's own change-order history should refine them. Second, the uplift is a funding reserve, not a construction target: management still builds to the base estimate, holds the difference centrally, and treats Section 20.7's reviews as the control on spending the reserve merely because it exists, a hazard the source paper itself flags.

Pre-mortem. A review exercise that assumes the project has already failed and asks each participant to write the history of the failure. It licenses dissent that ordinary review meetings suppress, and its output is a list of testable weaknesses.

20.7 *Pre-Mortems, Estimate Reviews, and the Audit Mindset*

Reference class forecasting fixes the budget line. The remaining defenses are procedural, and they are cheap relative to what they catch.

- **Pre-mortems.** Before approval, the team assumes the project failed and writes down why. The framing matters: asking ''what could go wrong?'' invites loyalty; asking ''it went wrong, what happened?'' invites the truth. The output feeds the sensitivity plan of chapter 12 with the adverse cases confirmation bias would have omitted.

- **Independent estimate reviews.** A review is independent only if it derives its own number before reading the advocate's (Example 20.2) and only if the reviewer does not report to the promoter. The Edinburgh case study below shows the form working in practice.

- **Governance: separate the estimator from the advocate.** The person whose career advances with approval must not own the estimate that justifies it. The

source paper's stronger versions: make forecasters carry the risk of their forecasts, subject them to outside review, and attach professional consequences to repeated flattery, measures whose intensity should rise with the organizational pressure, because that is where strategic misrepresentation lives.

This is also the place to say plainly what this book has been doing. Every chapter's *assessing outside recommendations* problems, the vendor sheets, the consultant memos, the confident AI analyses with one flaw, are training for exactly the failures cataloged here: anchors offered as answers, one-sided sensitivity, sunk costs dressed as arguments, benefits counted twice. The audit mindset those problems build, *rebuild the number before trusting it, and ask what the author needed you not to check*, is the individual-scale version of this section's governance, and it is the most durable thing a course in engineering economy can teach.

20.8 Case Study from the Literature

Chapter 1's case study read the overrun statistics and the UK uplift policy from Flyvbjerg (2006). This case study reads the same paper from the other end: the first recorded use of those uplifts in practice, an independent reviewer auditing a live business case, the audit mindset of Section 20.7 institutionalized.

Reference. Flyvbjerg (2006), Section on the Edinburgh Tram; open access at arXiv:1302.3642.

Example 20.6 — Case study: reviewing the Edinburgh tram business case

In October 2004, Ove Arup and Partners Scotland were appointed by the Scottish Parliament's tram committee to review the business case for Edinburgh Tram Line 2, which had been prepared on behalf of the promoter, Transport Initiatives Edinburgh. The submitted case estimated a base capital cost of £255 million plus a £64 million allowance for contingency and optimism bias, about 25%, for a total near £320 million. Arup, applying the UK Department for Transport's rail uplifts to the base cost, reported a 50th-percentile total of £357 million and an 80th-percentile total of £400 million, and noted that even these were likely low, because the scheme had not yet reached the outline business case stage at which the uplifts are calibrated. Arup also identified £46 million of specific omitted items (renewals and a revenue risk allowance), some 14.4% of the submitted total, and concluded that the promoter's optimism-bias allowance ''may have been underestimated.'' (a) Verify

Arup's two uplifted figures. (b) Compare the promoter's 25% allowance with the reference-class requirement. (c) Read the review as governance.

Solution.

Step (a): the uplifted totals.

Rail uplifts, 40% at the 50th percentile and 57% at the 80th, applied to the *base* cost:

$$255 \times 1.40 = \boxed{\text{£357 million}}, \qquad 255 \times 1.57 = 400.35 \approx \boxed{\text{£400 million}},$$

both matching the published review figures. Note the mechanics: the uplift replaces the promoter's own contingency, it is not stacked on top of it.

Step (b): the promoter's allowance against the class.

The promoter provided $64/255 = 25\%$ above base. The reference class says a rail promoter needs 40% merely to reach even odds, that is, $255 \times 0.40 =$ £102 million, and 57% (£145 million) for the 80% assurance a stand-alone public grant demands. The submitted allowance therefore bought *less than a coin flip* of staying within budget, while its rigor, a cost database, comparisons to other UK light-rail schemes, made it persuasive. That is this chapter in one sentence: an inside view can be simultaneously careful and biased, and only the outside view can tell.

Step (c): the governance reading.

Every safeguard of Section 20.7 appears in the episode. The reviewer was appointed by the funder's parliament, not the promoter; it derived its numbers from the reference class before reconciling with the submission, so it could not be anchored; and it put its conclusion, the allowance ''may have been underestimated,'' on the record before the funding decision. The review could not make the project's costs smaller; what it made smaller was the gap between what the decision-makers believed and what the evidence supported, which is all any forecast discipline can promise.

Insight. Reference class forecasting does not repair a biased estimate; it routes around it. Arup never argued with the promoter's quantity surveys, it re-based the total on

the outcome distribution of comparable projects, which no amount of inside-view care can shift. That is why the method survives strategic misrepresentation as well as honest optimism: it does not need the promoter's cooperation.

20.9 Chapter Summary

- The **planned-versus-actual record** is measured and stable: average constant-price cost overruns of 44.7% (rail), 33.8% (bridges and tunnels), and 20.4% (roads); rail demand 51.4% below forecast on average; no improvement over 70 years. The errors are **bias, not noise**.
- Two explanations survive testing: **optimism bias** (self-deception, strongest under low pressure) and **strategic misrepresentation** (deliberate shading, strongest where funds are competed for). They share a signature and need different cures.
- **Anchoring** makes reviews that start from the advocate's number non-independent: adjustments from an anchor stop too soon. Derive the review figure before reading the submission.
- **Sunk-cost escalation** funds failing projects because of what they have consumed. The forward rule of chapter 16 governs: compare completion cost against future value; the spent amount is identical in both futures. ''Finishing is right'' and ''starting was wrong'' can both be true.
- **Confirmation bias** turns sensitivity analysis into advocacy by testing only flattering directions. A table is evidence only when every input swings both ways (Section 12.3).
- **Reference class forecasting** corrects by the outside view: place the project in the outcome distribution of comparable completed projects. The UK **uplifts** (roads 15/32%, rail 40/57%, fixed links 23/55% at the 50th/80th percentiles) are that distribution turned into budget policy; the percentile is a risk choice, and lower uplifts must be earned by a demonstrated track record.
- **Pre-mortems**, **independent estimate reviews**, and **governance that separates estimator from advocate** are the procedural defenses; make accuracy pay and

flattery cost. The book's *assessing outside recommendations* problems train the same audit mindset at the individual scale.

20.10 Exercises

These exercises start with the overrun record and the uplift table applied to submitted estimates, then move to longer problems on the outside view and on sunk-cost escalation. The two audits that close the set are unusual in this book: every computation in them is correct, and the flaw to be named is anchoring in one memo and one-sided sensitivity in the other.

Warm-up and lecture practice

Exercise 20.1. Wadi Phosphate submits a 40-million-SAR estimate for a conveyor bridge across a wadi (fixed-links class). Set the budget at the 50% and 80% confidence levels, and state in one sentence who should use each figure.

Exercise 20.2. A promoter presents a benefit–cost ratio of 1.50 for a rail project and, separately, 1.50 for a road project. Using the record's averages (rail: costs $+44.7\%$, traffic -51.4%; roads: costs $+20.4\%$, traffic $+9.5\%$), estimate each realized ratio, assuming benefits scale with traffic. What do the two answers say about where forecast discipline matters most?

Quick checks (multiple choice)

Exercise 20.3. Which statement about optimism bias and strategic misrepresentation is **true**?

(A) They are the same thing under two names.

(B) Optimism bias is deliberate; strategic misrepresentation is innocent.

(C) They produce similar flattering forecasts, but one is self-deception cured by better method, and the other is deliberate shading cured by incentives and accountability.

(D) Both disappear when better forecasting models are adopted.

Exercise 20.4. A reference class forecast for a new concentrator is best described as:

(A) a more detailed bottom-up estimate of the concentrator's own costs
(B) the concentrator placed in the measured outcome distribution of comparable completed plants
(C) the average of the three lowest vendor quotations
(D) the base estimate plus a fixed 10% contingency

Exam-style problems

Exercise 20.5. Does the rail loop survive the outside view? Wadi Phosphate proposes a rail loop to the port: submitted capital cost 220 million SAR, net benefits 34 million SAR per year for 25 years, MARR 10%, $(P/A, 10\%, 25) = 9.0770$. (a) (*6 pts*) Set the 50%- and 80%-confidence budgets from the uplift table. (b) (*8 pts*) Compute the present worth at the submitted estimate and at both uplifted budgets (benefits unchanged). (c) (*4 pts*) State the verdict at each level in one sentence. (d) (*7 pts*) Find the break-even uplift, the cost growth that takes the present worth to zero, and interpret it against the rail class's 50th percentile.

Exercise 20.6. Escalation at the pellet plant. Najd Copper approved a pellet plant at 150 million SAR. It has consumed 120 million; finishing needs a further re-estimated 95 million; the completed plant's future net benefits are worth 130 million in present worth; abandonment salvage is 12 million. In the approval meeting a director argues: ''With 120 million invested, stopping now is unthinkable, we would be writing off two years of work.'' (a) (*4 pts*) Name the fallacy in the director's argument and the rule it violates. (b) (*8 pts*) Make the forward comparison and decide. (c) (*6 pts*) Find the completion cost at which the verdict flips. (d) (*7 pts*) Separately evaluate whether the *original approval* was vindicated, and state why the two evaluations must be kept apart.

Assessing outside recommendations

The two memos below fail behaviorally, not arithmetically: every computation in them is correct. Decide whether to accept each recommendation, name the bias, and correct the analysis. The flawed work is quoted exactly as received.

Exercise 20.7. Najd Copper commissions an ''independent review'' of the conveyor tunnel under the ridge (fixed-links class), submitted by its project team at

260 million SAR. The review memo reads:

> *"We have examined the project team's estimate of 260 million SAR and find its quantity take-offs thorough and its unit rates consistent with recent contracts. Applying a prudent margin of 8% for unforeseen conditions, we certify a review estimate of 280.8 million SAR. The estimate is sound."*

Accept or reject the review? Name the behavioral flaw, state what an independent review would have done, and produce the corrected figures.

Exercise 20.8. Wadi Phosphate's ore-sorting upgrade memo reports, at a MARR of 10%: base present worth $+6.0$ million SAR, and a sensitivity analysis: "throughput $+8\%$: PW $+9.8$M; sorter recovery $+2$ points: PW $+8.7$M; power tariff -10%: PW $+6.9$M. The present worth is positive in every case tested; the analysis is thorough and the project is approved for submission." (PW responds linearly: throughput swings PW by ± 3.8M per 8%, recovery by ± 2.7M per 2 points, tariff by ∓ 0.9M per 10%.) Accept or reject the memo? Name the bias, complete the analysis, and restate the conclusion.

References

Flyvbjerg, B. (2006). From Nobel Prize to project management: Getting risks right. *Project Management Journal*, 37(3), 5–15. Open access: `https://arxiv.org/abs/1302.3642`

Park, C. S. (2012). *Fundamentals of Engineering Economics* (3rd ed.). Pearson. [Cited throughout as "Park, 3rd ed."]