The Decision Architecture

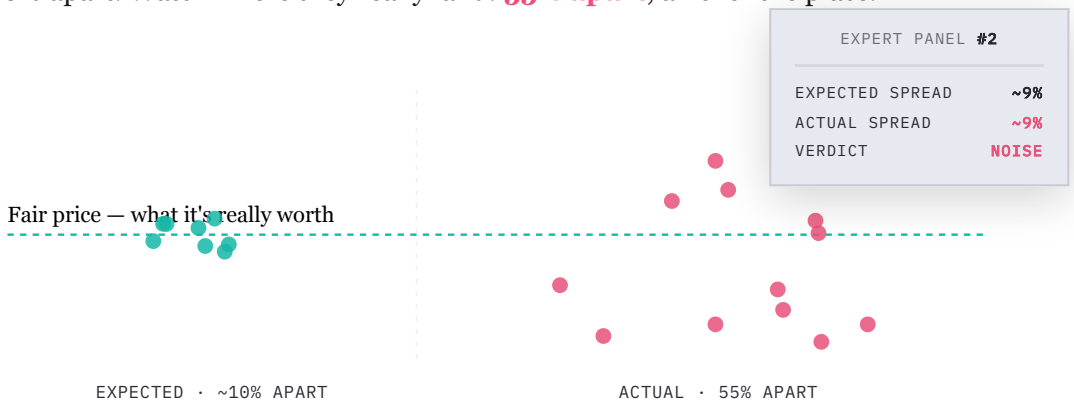
James E. Coleman

STRATEGIC DESIGN · WHITE PAPER SERIES · 2025

ABSTRACT *Closing the intelligence gap is wasted if the decisions that consume that intelligence are themselves poorly designed. Better dashboards do not produce better choices — designed judgment does. Organizations treat decision-making as individual talent; it is an architecture that can be engineered to cut bias and noise and to raise calibration.*

FIG. 1 — WATCH THE SAME RISK GET PRICED ACTUAL GAP: 9%

Each dot is one expert pricing the **exact same** risk. They expected to land **close together** — about 10% apart. Watch where they really land: **55% apart**, all over the place.



THE NOISE WE REFUSE TO SEE



55%

Two experts pricing the **same** risk land about 55% apart — where leaders expected roughly 10%.[1][2]



9 in 10

Nine in ten large transport projects run over budget — the planning fallacy, uncorrected.[3]



94%

Across 136 studies, a **simple formula** matched or beat the human expert 94% of the time.[4]



<1%

And fewer than 1% of firms ever run a **noise audit** of their own decisions.

§ 01 The Judgment Gap

Intelligence terminates *in a decision.*

WP 1.0 argued that advantage flows from the capacity to convert data into decision-quality intelligence at speed. But intelligence is never the terminal product of that pipeline — a decision is. Every dashboard, model, forecast, and briefing exists to be consumed by an act of judgment, and it is that act, not the input feeding it, that commits capital, people, and reputation. Organizations have poured a generation of investment into the *inputs* of judgment and almost none into judgment itself.

The asymmetry is stark. A firm will spend millions on a data warehouse and nothing on the protocol by which that warehouse is actually read into a choice. It treats the decision as a personal attribute of senior people — talent, instinct, pattern-recognition earned over decades — rather than as a process that can be specified, measured, and improved. Yet the research that first mapped the limits of human judgment showed those limits to be *systematic*: intuitive judgment runs on a small set of heuristics that are efficient but predictably fallible, not on the dispassionate weighing of evidence we imagine.[5]

We name the resulting shortfall the judgment gap: the distance between the quality of intelligence an organization can produce and the quality of the decisions it actually makes with that intelligence. A world-class intelligence function feeding an undesigned decision process yields nothing more than world-class inputs to an unmeasured, unaccountable act. This is where advantage leaks out — not upstream in the data, but at the last mile, in the move from knowing to choosing. Better dashboards do not close it. Designed judgment does.



NASA's Mission Operations Control Room during Apollo 13 — decision-making under uncertainty, fully instrumented.

Photo: NASA · Public domain

> *"Most organizations can tell you to two decimal places what their data cost. Almost none can tell you whether their decisions are any good. The input is instrumented; the judgment that spends it is not."*
>
> — JAMES E. COLEMAN, DECISION ARCHITECTURE FRAMEWORK

The remedy is not to replace human judgment — the evidence for doing so is weaker than it is often made to sound — but to treat judgment as an engineering surface. The error it produces can be decomposed, its sources named, and each source attacked with a specific technique. The first step is to see clearly what kind of error we are fighting.

Two error types. *Organizations fight only one.*

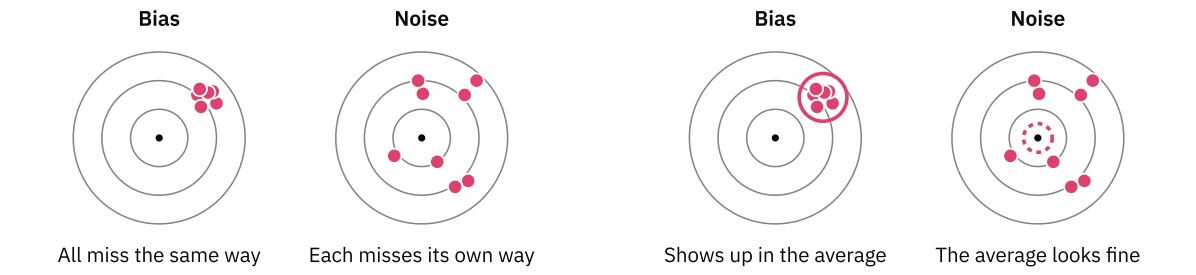
Error in judgment comes in two forms, and most organizations are fluent in only one of them. Bias is shared, directional error — the whole team leaning the same wrong way at once. It is the signature of the heuristics: anchoring on an initial number, over-weighting vivid recent events, systematic over-optimism about one's own plans.[5] Prospect theory sharpened the point by showing that even our risk preferences are biased in patterned ways — losses loom larger than equivalent gains, and the mere framing of an option as a loss or a gain flips the choice.[6]

Noise is the error organizations rarely name: unwanted variability in judgments that are supposed to be identical. If bias is a cluster of shots landing off-centre together, noise is a cluster scattered across the target — equally wrong, but in every direction at once. When a large insurer audited its own underwriters, executives guessed that two professionals pricing the same risk would differ by about 10%. The median difference was 55% — five times larger than the number the business was unknowingly built on.[1][2] The same scatter turns up in credit decisions, medical diagnoses, hiring, and criminal sentencing.

Noise is under-recognized because it is structurally invisible. Bias eventually shows up in the aggregate — a forecast that is always too high can be caught. Noise hides, because an organization almost never sees two judgments of the *identical* case placed side by side; each decision is made once, by one person, and the variability never surfaces. It is nonetheless large, costly, and — crucially — *measurable*, through a noise audit that most firms have simply never thought to run. Bias and noise are independent problems that demand different fixes.

FIG. 4.1 — BIAS VS. NOISE

STATE 1 OF 4



WHAT EACH ONE IS

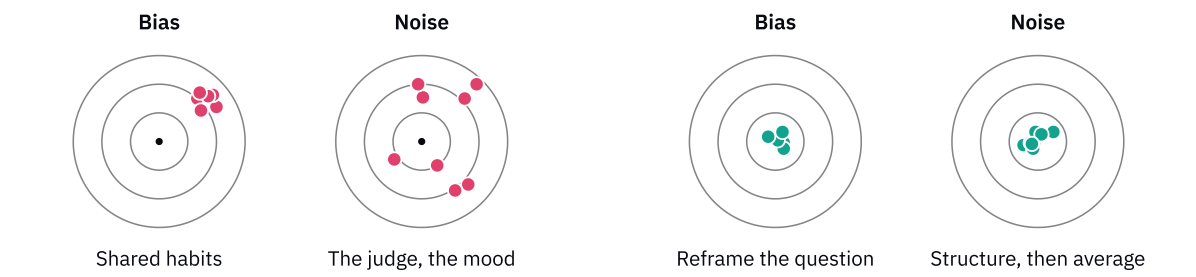
Definition

Bias: shared, directional error — judgments miss in the same direction. **Noise:** unwanted scatter — judgments of the same case miss in different directions.

HOW EACH ONE IS FOUND

Visibility

Bias: detectable in the average outcome over time. **Noise:** invisible without deliberately comparing independent judgments of identical cases — so it usually goes unmeasured.



WHERE EACH ONE COMES FROM

Typical source

Bias: heuristics, incentives, motivated reasoning shared across a group. **Noise:** idiosyncratic differences between judges, plus mood, order, and fatigue within a single judge.

WHAT EACH ONE NEEDS

What reduces it

Bias: reframing, outside-view checks, debiasing the question. **Noise:** structure, aggregation of independent estimates, and replacing or disciplining discretion with rules.

A diagram of the relation, not a chart of data. Each dot is one judgment of the same case, and the centre is the right answer. In the last state the green dots show each error after its remedy.

§ 03 Decision Hygiene & Forecasting Discipline

Discipline the judgment. *Don't abolish it.*

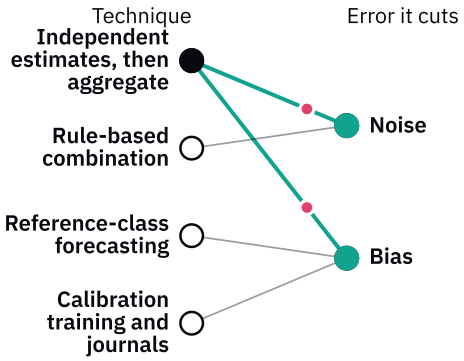
The practices that reduce decision error behave like *hygiene*: you apply them without knowing which specific error they prevented on any given occasion, the way hand-washing stops an infection you never identify. The foundational move is to gather estimates independently before any discussion. The moment a group hears the first confident voice, judgments cascade and anchor; collecting each person's number in private first preserves the diversity that makes aggregation powerful, and the simple average of independent estimates cancels a large share of noise for free.

The most uncomfortable evidence concerns the judgments themselves. Across decades of head-to-head comparisons, a simple statistical rule matched or outperformed expert clinical judgment in 128 of 136 studies — the expert won outright in only eight.[4] The lesson is not that expertise is worthless but that unaided, case-by-case discretion is *noisy*: experts disagree with each other, and with themselves on different days.[7] Decision hygiene decomposes a judgment into components, scores each against a reference where one exists, and combines them mechanically — borrowing the rule's consistency while keeping human judgment for the inputs machines cannot supply.

Reference-class forecasting attacks bias from the outside. Rather than build a plan from its own optimistic internals, it asks how comparable past efforts actually turned out — and the base rate is sobering: roughly nine in ten large transport projects run over budget, with rail averaging overruns near 45%.[3] The outside view corrects the inside story. Finally, calibration training, decision journals, and scored forecasts close the feedback loop: in a multi-year geopolitical forecasting tournament, participants given modest probabilistic training and consistent scoring became measurably more accurate, and the best — the "superforecasters" — did so by tracking, grading, and revising their own predictions as a matter of routine.[8]

FIG. 4.2 — TECHNIQUE → ERROR REDUCED

STATE 1 OF 4



CUTS NOISE AND CASCADE BIAS

Independent estimates, then aggregate

Cuts **noise** and cascade **bias**. Collecting judgments privately before discussion preserves diversity; averaging independent estimates cancels random error.

CUTS NOISE

Mechanical / rule-based combination

Cuts **noise**. Decomposing a decision and combining component scores by rule removes the within- and between-judge variability that defeats unaided discretion.

CUTS BIAS

Reference-class forecasting

Cuts **bias** (optimism, the planning fallacy). Anchoring a forecast to the outcomes of a comparable class replaces the inside view with an empirical base rate.

CUTS BIAS, IMPROVES CALIBRATION

Calibration training + decision journals

Cuts **bias** and improves **calibration**. Recording rationale and probabilities, then scoring outcomes, supplies the feedback that turns experience into accuracy rather than confidence.

A diagram of the relation, not a chart of data. Each line joins a technique to the kind of error it reduces.

§ 04 Designing the Decision Process

Judgment is *an architecture, not a talent.*

The techniques of §03 are not a menu for gifted individuals to adopt privately. They are the components of an organizational architecture — and the shift that matters is to stop asking "who is our best decision-maker?" and start asking "what is our decision process?" The first design act is to classify decisions along the axes that determine which method fits: reversible or irreversible, frequent or one-off, high-noise or low-noise. A frequent, reversible, high-noise decision — pricing, triage, screening — should be routed to a rule or a decision aid. A rare, irreversible, bias-prone decision — an acquisition, a market entry — should be routed to independent estimates, an outside view, and a structured premortem.

Architecture also means owners and feedback loops. Every consequential decision carries a named owner, a recorded rationale, and an explicit probabilistic forecast of its outcome, filed at the moment of choice and reviewed on a schedule. That discipline produces the one asset almost every organization lacks: *data on its own decision quality*. Without the recorded forecast there is nothing to score, and without scoring there is no calibration — only the comfortable narrative that hindsight always supplies. The decision journal is the instrument that makes judgment, at last, measurable.

WP 1.0 named "decision architecture" as one of the domains where intelligence must terminate; this paper specifies the engineered system that domain requires. And it closes the series' central loop. Intelligence, translational capacity, and trust are accumulated assets — they compound when invested in deliberately and decay when neglected. *Calibrated judgment is the same kind of asset.* A disciplined decision archive is not overhead; it is capital that appreciates, the mechanism by which an organization learns to choose better over time rather than merely to know more. The firms that will define the next era are not those with the most intelligence, but those that have designed the architecture to spend it well.

> *"A decision made once and forgotten teaches nothing. A decision made with a recorded forecast, a named owner, and a scheduled review becomes an entry in an appreciating asset — the organization's accumulating stock of calibrated judgment."*
>
> — JAMES E. COLEMAN, WP 4.0

↳ The series thruline

The gap between knowledge and action is an infrastructure problem, not a resource problem — crossed by systems designed around the people who do the crossing. Here the gap runs from information to judgment. The same move recurs across the series: build the system the people who do the crossing currently lack.

§ R References — Primary Sources

Sources *& notes.*

Every empirical claim in this paper is cited to a primary or seminal source below, and each link was verified to resolve. The underwriting-noise figure (55% vs. an expected ~10%), the planning-fallacy base rate (≈9 in 10 projects over budget), and the clinical-versus-actuarial tally (128 of 136 studies) are drawn directly from the cited studies. The single hero tile marked modeled estimate · illustrative — the share of firms that ever run a noise audit — is labeled as such on the page: the literature establishes that noise audits are rare, but does not fix a precise prevalence.

1. [1]Kahneman D, Rosenfield AM, Gandhi L, Blaser T. “Noise: How to Overcome the High, Hidden Cost of Inconsistent Decision Making.” *Harvard Business Review* 2016; Oct. Reports the underwriting noise audit — a median 55% difference between underwriters pricing the same risk, against an expected ~10% — and a 43% difference among claims adjusters. hbr.org/2016/10/noise.

2. [2]Kahneman D, Sibony O, Sunstein CR. *Noise: A Flaw in Human Judgment*. Little, Brown Spark, 2021. Book-length treatment distinguishing bias from noise and setting out the noise audit and decision-hygiene practices. hachettebook-group.com.

3. [3]Flyvbjerg B, Holm MS, Buhl S. "Underestimating Costs in Public Works Projects: Error or Lie?" *Journal of the American Planning Association* 2002; 68(3): 279–295. Cost underestimation found in ~9 of 10 large transport projects; actual rail costs average ~45% above estimate. Empirical basis for reference-class forecasting. doi:10.1080/01944360208976273.

4. [4]Grove WM, Meehl PE. "Clinical Versus Mechanical Prediction: A Meta-Analysis." *Psychological Assessment* 2000; 12(1): 19–30. Across 136 studies, mechanical prediction matched or exceeded clinical judgment in 128; clinical judgment was clearly superior in only 8. doi:10.1037/1040-3590.12.1.19.

5. [5]Tversky A, Kahneman D. "Judgment under Uncertainty: Heuristics and Biases." *Science* 1974; 185(4157): 1124–1131. Foundational account of the availability, representativeness, and anchoring heuristics and their systematic biases. doi:10.1126/science.185.4157.1124.

6. [6]Kahneman D, Tversky A. "Prospect Theory: An Analysis of Decision under Risk." *Econometrica* 1979; 47(2): 263–291. Establishes loss aversion and framing effects — biases in risk preference itself. doi:10.2307/1914185.

7. [7]Dawes RM, Faust D, Meehl PE. "Clinical Versus Actuarial Judgment." *Science* 1989; 243(4899): 1668–1674. Reviews the consistent advantage of simple actuarial models over unaided expert judgment, attributing much of the gap to the noisiness of human discretion. doi:10.1126/science.2648573.

8. [8]Mellers B, Ungar L, Baron J, et al. "Psychological Strategies for Winning a Geopolitical Forecasting Tournament." *Psychological Science* 2014; 25(5): 1106–1115. Good Judgment Project evidence that probabilistic training, teaming, and tracking measurably improve forecasting accuracy — the empirical core of calibration discipline. doi:10.1177/0956797614524255.

SERIES CLOSING STATEMENT

"Intelligence, translation, trust — and the judgment that spends them — are all accumulated assets. They compound when invested in deliberately and decay when neglected. The organizations that will define the next era are not those with the most intelligence, but those that have designed the architecture to spend it well — treating

the decision itself as a system to be built, maintained with intention, and evaluated against outcomes."

— JAMES E. COLEMAN — SERIES EDITION 2025