

RED BEAR ARTIFICIAL INTELLIGENCE

ABU DHABI, UNITED ARAB EMIRATES

Decision Impact Analysis

What a decision intelligence dataset can honestly tell you, and what it will be sold as

A structured record of how decisions were actually made supports a great deal of useful analysis and rather less causal inference than it will be sold as, and the useful part is worth having on its own. This paper sets out the four kinds of analysis a decision dataset supports, the four reasons a causal claim from it usually fails, and how Red Bear: Enterprise Decision Intelligence produces the dataset while refusing the claims it cannot support.

Empower. Trust. Verify.

White paper EDI-WP-06 · Red Bear white paper series · Version 2.0

26 September 2026

Summary

Every organisation measures its outcomes, and almost none can tell you which decisions produced them. Revenue, margin, churn, incident rate and time to market are instrumented to three decimal places. The choices that moved them are not instrumented at all. They are attributed instead, to strategy, to the market, to execution, to the quarter being a difficult one, by people who already know how it turned out.

This paper is about what would change if decisions themselves were data: classified consistently, recorded with the reasoning current at the time, and joined to what happened afterwards. It argues that such a dataset would be genuinely valuable, that most of the value is descriptive rather than causal, and that the causal claims which will inevitably be made on top of it deserve more scepticism than they usually receive. The audience for this paper has been sold analytics before. It is written accordingly.

If the unit existed, four kinds of analysis become possible, and they are not equally achievable. Description is immediate and reliable. Pattern recognition is straightforward and frequently mistaken for insight. Prediction is feasible where the class is high-volume and the outcome is measured quickly. Causal inference is hard, mostly unavailable from observational data, and is the thing most likely to be promised. The decisions that reach a register are not a random sample; the conditions that produced the decision usually also affect the outcome; the cases that failed badly enough disappear from the record; and for any single decision you never observe what would have happened had the other option been chosen.

The honest near-term value is therefore not a causal engine. It is four descriptive things, each actionable without any causal claim: a baseline of how a class is actually decided; the places where recorded reasoning systematically failed to anticipate the result; the policies overridden so often that they are not controls; and the decisions slow enough to be costing money. And a fifth, the strongest argument in the paper: where several people assess the same case independently, the spread between their judgements, noise, is directly measurable, needs no counterfactual, and one of them is wrong whatever happens next.

Red Bear: Enterprise Decision Intelligence (formerly GuardrAll) produces the dataset. It shows what a decision touched and how each object performed since, against an expected trajectory fixed before the decision, with confounders at equal weight, and it never reports deviation as what the decision caused.

WHO SHOULD READ THIS

- Chief financial officers and heads of finance analytics
- Chief data and analytics officers
- Chief risk officers asked to evidence the effect of a control
- Anyone who has been shown a decision analytics dashboard and asked how the effect was identified

WHAT YOU WILL BE ABLE TO DO AFTERWARDS

- Place any proposed analysis on the four rows, and say what it needs and what it cannot deliver
- Name the four reasons a causal claim from observational decision data usually fails
- Run a noise audit on one decision class in a few days without a platform
- Ask a vendor, including this one, how they identified the effect

1 What it is

Consider how a decision leaves a trace today. A pricing exception is agreed in a call. It appears in the CRM as a changed field, in the ERP as an order at a non-standard price, and in the month-end pack as a variance. Three systems hold three fragments. None holds the decision. Nobody recorded which options were considered, what was assumed about the customer's alternatives, who disagreed, or what the decision-maker expected to happen. Six months later the account is retained at a lower margin, and that result is available to be attributed to whatever narrative is convenient.

This is not an oversight. Systems of record were designed to record obligations and balances, because those must be reconciled. Judgement has no reconciliation requirement, so it was never given a schema. Without a decision as a unit, every question about decision-making is answered by proxy: how long we take to decide, from ticket ages, which measure queueing; how often we override our own controls, from audit exceptions, which are a sample of the ones that were found; how good our judgements are, from outcomes, which conflates decision quality with luck, the confusion Duke names *resulting* (Duke, *Thinking in Bets*, chapter 1).

Decision impact analysis is what becomes possible when the unit exists: a consistent record of decisions, their class, their reasoning, the policy in force, who decided, how long it took, the business objects they were about or affected, and what those objects did afterwards. Assume, for the moment, that the dataset exists. What can be done with it? Four things, in ascending order of difficulty and descending order of reliability.

Analysis	What it answers	What it needs	What it cannot deliver
Description	How this class is actually decided: by whom, how fast, on what evidence, how often escalated, how often overridden	Consistent classification and complete capture; no outcomes required	Any statement about whether the way it is made is the right way
Pattern recognition	Which recorded features move together: reasoning completeness against reversal rate, speed against later amendment	Enough volume per class for the pattern to be distinguishable from noise	Direction: a correlation between two recorded features says nothing about which, if either, produces the other
Prediction	Given what is recorded at the moment of decision, the likely outcome and how confident to be	High volume, a stable class, outcomes that resolve quickly enough to train on	Any account of why; a model that predicts well under one policy typically fails when the policy changes
Causal inference	Whether deciding differently would have produced a different result	A credible source of variation unrelated to the outcome: a designed experiment or a genuine natural one	Almost everything, from observational data alone

Two observations are worth more than the table. The value does not increase down the rows: description is the least glamorous row and by a wide margin the most likely to change a decision next quarter. And the requirements column is where most programmes fail: not at the modelling, which everyone is confident about, but at the classification and the outcome horizon, which are unglamorous, contested, and have to be settled before any data is collected.

Key points

- Decisions are the missing unit of analysis; outcomes are analysed and decisions inferred, which is the wrong way round
- Description, pattern, prediction, cause: reliability falls as ambition rises

- Register first, discipline second, analysis third; the order is not negotiable

2 Why it matters

A structured decision dataset will, reliably, produce findings of the following shape: decisions in class X, taken under condition Y, were followed by materially worse outcomes than decisions in class X taken otherwise. The finding will be arithmetically correct. The temptation to act on it as though Y caused the worse outcome is very strong, and in most cases it is not supported. Four reasons, and they compound.

Reason	What it does to the finding
Confounding	The condition that produced the decision usually also affects the result. Capital approvals escalated to the executive committee show worse returns than those settled at divisional level; the more likely reading is that difficult, novel, high-variance proposals are the ones that get escalated. Controlling for what you recorded does not fix this, because the confounder is usually the thing nobody thought to record
Selection into the record	Decisions that reach a register are the ones consequential, formal or visible enough to be captured, and early in a register's life they are the decisions made by the teams who complied first. Neither is a random sample
Survivorship	Records of decisions that went badly enough have a habit of not being there: the supplier is no longer a supplier, the product line was discontinued, the executive who approved it has gone. Surviving records flatter whatever practice they followed
The counterfactual you never observe	For any given decision you observe one branch. The comparison that would establish causation, the same decision, the same conditions, the other choice, is unavailable in principle. No amount of collection resolves this

Where a causal claim can honestly be made. A policy threshold at a round number creates a discontinuity: decisions just above and just below a limit are similar in every respect except the governance treatment, so comparing across the boundary isolates the effect of the treatment. A staged rollout of a new approval route across regions is a natural experiment, provided the regions were not chosen for their performance. A deliberate randomisation, routing a proportion of a high-volume class to a different review path, gives the cleanest answer available and is practical for classes that run at thousands a month. All three have to be planned. None can be recovered by mining a dataset afterwards, and an analysis that arrives at a causal claim by searching until one appears has found a coincidence with good presentation. When a threshold is being set or a route deployed, somebody should be asked whether it can be arranged to be analysable. The cost of asking is an afternoon. The cost of not asking is that the question becomes unanswerable for ever.

A worked example of how hard this is comes from a study often cited in support of decision process. Lovallo and Sibony asked executives to describe 1,048 decisions, the analysis performed, the process followed and how it turned out. Process accounted for around six times as much of the variation in outcome as the quality of the analysis. It is a striking result and it points in a sensible direction. It is also self-reported and retrospective: the same executives rated the process and the outcome, knowing both, with no counterfactual and no independent measurement. That is precisely the design problem this paper describes. A register that records the process at the time, before the outcome is known, removes one of its largest weaknesses, which is the reason to build one (Lovallo and Sibony, *The case for behavioral strategy*, McKinsey Quarterly, 2010).

The value that survives the scepticism. Strip out every claim that requires causal identification and a substantial amount remains, each descriptive, each measurable without a model, each with a defensible action attached. A baseline of how a class is actually decided, which most organisations have never seen for any class, and which is not

contestable. Where reasoning fails to anticipate the result: if a class routinely justifies itself on an assumption that later proves wrong, that is a pattern in the record, not a judgement about an individual, and a prompt to change the questions asked at capture. Policies that are not controls: a policy overridden in most of the cases it applies to is generating paperwork and a false assurance, and the remedy is a management decision. Decisions slow enough to be costing something: the waiting is measured; the assumption that converts waiting into money should be stated rather than buried. And noise. Where two or more people assess the same case independently before they confer, the disagreement between them is measurable directly, needs no outcome and no counterfactual: if judgements meant to be interchangeable are not, the organisation has a consistency problem that exists whether or not the outcomes turn out well (Kahneman, Sibony and Sunstein, *Noise*, part 1). Bain's four dimensions, quality, speed, yield and effort, frame the descriptive layer; three of the four are observable from a register without any causal claim, and only quality requires judgement (Blenko, Mankins and Rogers, *Decide & Deliver*, chapter 1).

3 The challenges

The analysis is only available if the dataset has particular properties, and they have to be designed in. Retrofitting any of them is expensive and retrofitting the third is impossible.

Challenge	Why it is hard	What goes wrong without a method
Consistent classification	A maintained catalogue of classes, with each decision assigned by a rule rather than by the preference of whoever recorded it	Every comparison is between groups assembled differently; apparent differences between classes are artefacts of labelling
Outcomes at a defined horizon	The review point is set when the class is defined and applied to every decision in it	Outcomes are reviewed when someone remembers, disproportionately when something went wrong: the sample is selected on the dependent variable
Reasoning captured at the time	Options, assumptions, expected result and dissent recorded before anyone knows how it turned out	Reasoning reconstructed later explains the outcome rather than predicting it, and cannot be used to assess process
Policy and context as they stood	The rule and limits in force on the day, versioned	Decisions are assessed against today's policy, which makes historical practice look worse or better than it was
Coverage that is not self-selecting	Capture where decisions already happen, so reaching the record is not an act of discretion by the decider	The register fills with the decisions people were comfortable recording, and every analysis inherits that filter
Objects and measures with provenance	The business objects a decision acts on, with their identifiers in each source system, and a time series of performance measures defined to the tenant's own definitions	Impact is asserted against a number nobody can trace to a source and a read time

Most organisations have none of these, which is the practical conclusion of this paper. The sequence is register first, discipline second, analysis third. Reversing it produces a modelling exercise on data that cannot support the questions being asked of it.

4 How Red Bear: Enterprise Decision Intelligence helps

The product's contribution to this argument is narrow and should be described narrowly: it produces the dataset, and it refuses the claims the dataset cannot support.

Challenge	What Enterprise Decision Intelligence does	Requirements
Consistent classification	A maintained catalogue of classes, which customers extend; each candidate classified with a confidence, and held for a person below the threshold; the class sets the evidence, memory and context standards	FR-REC-01, FR-REC-06
Outcomes at a defined horizon	Outcome tracking at the class's interval, from a connected system reporting against the decision identifier or the accountable owner's review; an outcome the product guessed is not evidence	FR-OUT-01, RBP FR-DL-01
Reasoning captured at the time	The four Cartesian quadrant answers written by the requester at the point of decision; confidence per decision and per assumption, immutable once written; dissent and pre-mortem attributed and timestamped	FR-CAP-12, FR-REC-12, FR-CAP-11
Policy and context as they stood	Every evaluation records the policy version it used; audit replay reconstructs the record, evidence, policy, dial and the context the decider saw, never re-rendered	FR-DEL-14, FR-REC-09
Coverage that is not self-selecting	Capture through productivity, messaging and enterprise-application connectors, voice intake and the portal, all terminating in one candidate contract; harvested history seeds a baseline, marked harvested and excluded from quality scoring	FR-CAP-07, FR-CAP-03, FR-CAP-09
Objects and measures with provenance	The subject and measure model holds the objects decisions act on with their identifiers in each source system and a time series of the tenant's own measures; a measure without provenance is not admitted; every decision records its objects as *about* or *affected* with a confidence	FR-OUT-09
Description	The five diagnostics computed continuously per class with trend: cycle time from two timestamps, noise as the spread between independent judgements, the exception ratio read rather than tallied, the reconstruction rate, and the standing unmanaged-risk question; no invented return figure sits between baseline and indicators	FR-OUT-07
Policies that are not controls	Override telemetry: which policies, how often, with what reasons, presented as a question about the policy	FR-DEL-18
Decisions slow enough to cost	Bottleneck and latency analysis: where decisions wait, for whom, against target, and what the delay costs, with the assumption stated	FR-RTE-05
Impact without causation	The impact explorer shows what a decision touched and how each object performed since, against an expected trajectory computed from history before the decision only, stored with its method and never recomputed; deviation reported with the measure's own volatility and never as what the decision caused; confounders shown at equal weight; below the attribution threshold no figure is shown and the reason is stated	FR-OUT-10, FR-OUT-03
Prediction, bounded	Decision impact prediction forecasts the likely effect of a pending decision from precedent and context, always with its confidence and the count of precedents; below the configured count it shows the precedents and declines to forecast	FR-SUP-03
Analysable by design	Policy versions, dial moves, routing changes and sandbox replays are recorded with the decisions they affected, so a threshold discontinuity or a staged rollout can be planned and later analysed	FR-DEL-14, FR-DEL-16

Two capabilities bear directly on this paper and should be described precisely, because their names invite more than they do. Decision impact prediction is prediction in the sense of the third row: at its best in a high-volume class whose outcomes resolve quickly, offering no account of why, degrading when the policy changes. Legacy harvesting seeds a baseline so that description does not wait two quarters for the register to fill; a harvested record is reconstructed

rather than contemporaneous, is marked as such, and is sound evidence of how a class was handled and how long it took, not of the quality of its reasoning, because the reasoning was written by people who already knew the answer.

Design rule

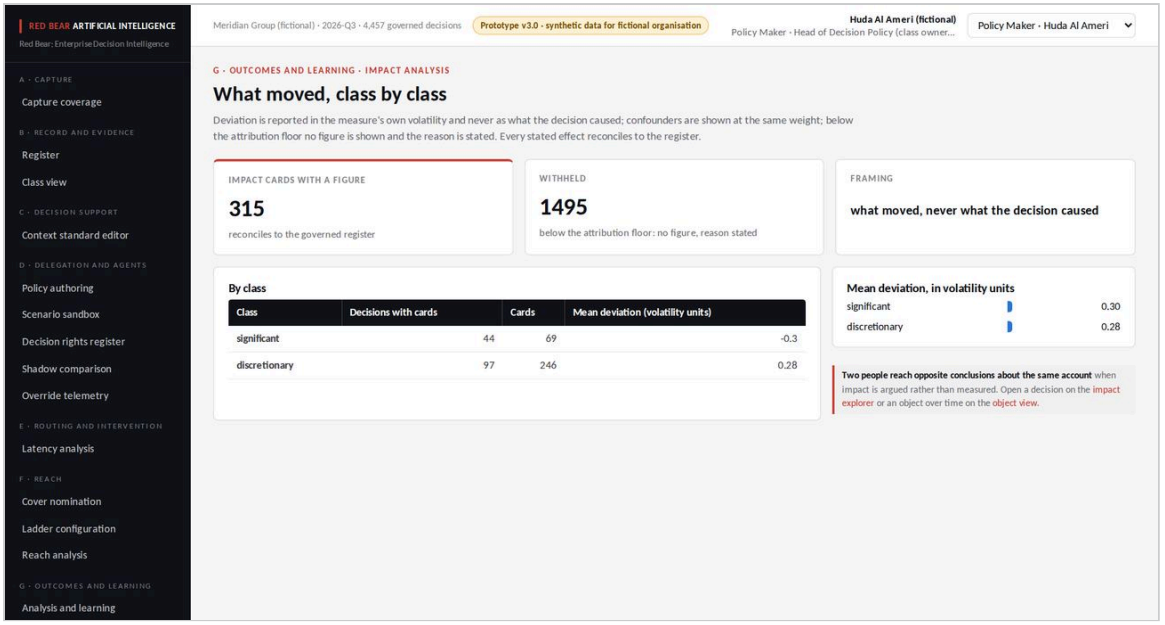
The product performs decision impact analysis; it does not establish that deciding one way rather than another caused a business outcome. A register makes description reliable, pattern recognition possible and noise measurable. It makes credible causal work feasible to design. It does not make causal claims safe to assert from observation, and any vendor who says otherwise, including this one, should be asked how they identified the effect.

5 How it is used

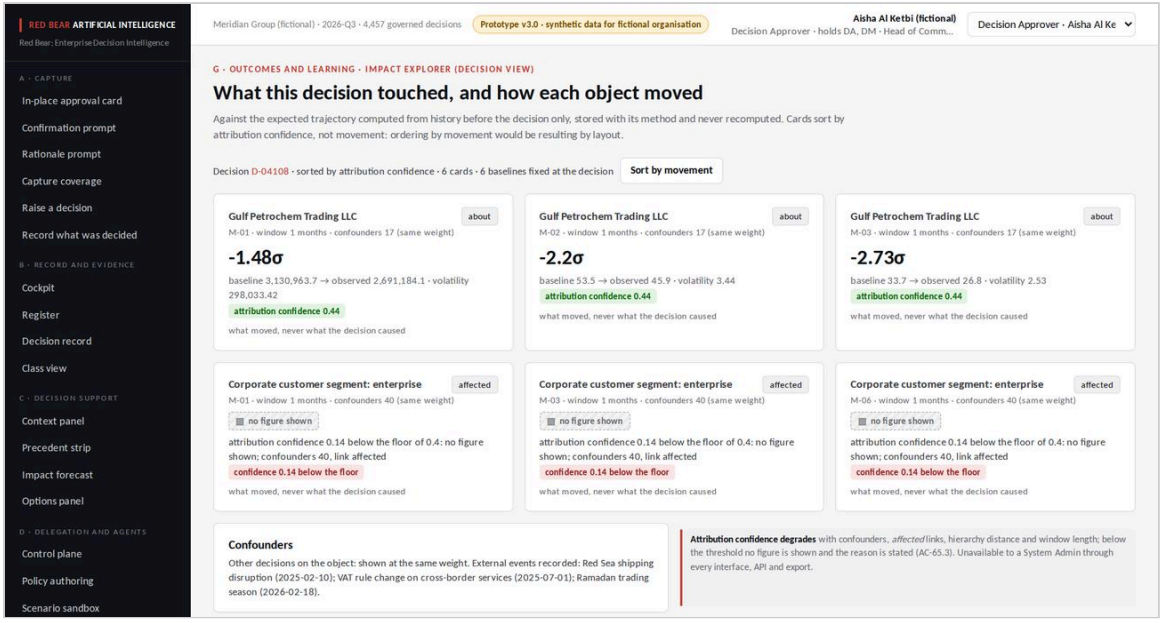
Step	Who	What happens
1 Pick one class	Sponsor, class owner	High-volume, consequential and currently invisible: pricing exceptions, credit limit changes, supplier selection below the board threshold, scope changes on delivery
2 Define the horizon first	Class owner	The review interval and what counts as a good outcome at it, before looking at any data; the argument this provokes is itself informative
3 Run a noise audit	Class owner, assessors	Six to ten recent cases, given to several qualified assessors independently; record the spread before anyone confers; a few days, no platform
4 Baseline	Sponsor	Decision speed, bottleneck cost and override rate measured on the existing process before anything is switched on; harvested history seeds description
5 Capture properly	Decision Maker	Reasoning at the time, policy as it stood, objects named, outcome at the horizon; two quarters before expecting analysis of any kind
6 Read the descriptive layer	Class owner, CFO	The five diagnostics, override telemetry and latency: each with a management action and no causal claim
7 Design the causal question	Class owner, analyst	Where a threshold is being set or a route deployed, arrange for it to be analysable; hold whoever proposes the first causal finding to the standard above

If the noise audit shows a narrow spread, the case for instrumenting the class is weaker and you have learned that cheaply. If it is wide, you have a measured finding that needs no causal defence, and a reason to continue. If a commercial platform is the right way to do the capture, that is a procurement question, and it comes after these steps rather than instead of them.

6 The user experience



S-44 Impact analysis: S-44 Impact analysis. The actual effect of a class of decisions on cost, revenue, risk and time, every stated effect reconciling to the register, beside the five diagnostics and their trend. Prototype render on synthetic data for fictional entities.



S-45 Impact explorer (decision view): S-45 Impact explorer (decision view). What one decision touched, each object's measures since against the expected trajectory fixed before the decision, confounders at equal weight, attribution confidence, and no figure where confidence is below the threshold. Prototype render on synthetic data for fictional entities.

7 Where the data comes from, and how it is checked

Inputs to impact analysis

Input	Source	Check
Decisions and their reasoning	Every capture route, terminating in one candidate; the point-of-decision prompt	Confirm-first; deduplicated on the decision; classified with a confidence (FR-CAP-12, FR-CAP-07, FR-REC-01)

Input	Source	Check
Objects the decision acts on	Named on the record as *about* or *affected*, with a confidence; identifiers resolved per source system	Nesting and identifiers held in the subject model (FR-OUT-09)
Performance measures	Time series read from connected systems, defined to the tenant's definitions	Every observation carries its source system and read time; a measure without provenance is not admitted (FR-OUT-09, RBP FR-PV-01)
Expected trajectory	Computed from history before the decision only	Stored with its method and never recomputed (FR-OUT-10)
Confounders	Other decisions on the object in the window; recorded external events	Shown at the same weight as the decision (FR-OUT-10)
Outcomes	Connected systems against the decision identifier; the owner's review at the horizon	An outcome the product guessed is not evidence (FR-OUT-01)
Historic decisions	Legacy harvesting, once per engagement	Marked harvested; excluded from quality scoring, value baselines and metering (FR-CAP-09)

How it is checked

- **Connectors are read-only and consented.** Measures and outcomes are read under least privilege granted by the customer administrator; the product writes nothing into a source system's numbers (RBP FR-CN-01, RBP FR-CN-03).
- **Provenance travels with every number.** Source system and read time on every observation; evidence as immutable snapshots by digest (FR-OUT-09, FR-REC-12).
- **Attribution degrades honestly.** Confidence falls with confounders, *affected* links, hierarchy distance and window length; below the threshold no figure is shown and the reason is stated; cards sort by confidence, not movement (FR-OUT-10).
- **The AI proposes; a person confirms.** A suggested class, a proposed precedent and a forecast are proposals; the forecast carries its confidence and precedent count and declines below the configured count (FR-REC-01, FR-SUP-03).
- **Reconciliation is required.** Every stated effect reconciles to the register; no invented return figure sits between the baseline and the indicators (FR-OUT-03, FR-OUT-07).
- **Patterns, never individuals.** Rationale and quality patterns are computed across classes and sliced by region and type, never by name; the impact explorer is unavailable to a System Admin through every interface (FR-OUT-08, FR-OUT-10, RBP FR-PR-05).

8 Benefits realised

Every measure is calculated from the organisation's own register against a baseline established before anything was switched on (FR-PLT-05).

Benefit	How it shows up	Measured by
A baseline nobody has seen	Who decides, how long, on what evidence, how often escalated, how far practice varies between teams	The five diagnostics per class with trend

Benefit	How it shows up	Measured by
Noise made visible	The spread between independent judgements on the same case	Noise per class; the cheapest finding in the paper
Policies that govern	Overridden rules changed or enforced	Override frequency per policy, falling or explained
Waiting that is priced	Where decisions wait and what the delay costs, with the assumption stated	Bottleneck cost per class
Reasoning that anticipates	Assumptions that repeatedly prove optimistic are changed at capture	Assumption accuracy over time, per class
Impact that reconciles	Effects on cost, revenue, risk and time trace to register records	Stated effects reconciled to the register: all of them

9 Risks managed

The strongest objection is not about statistics. It is about proportionality: this is a large amount of instrumentation in pursuit of insight an experienced executive team already holds intuitively. They know which approvals are slow, which policies are ignored, which of their people decide well. Part of this is correct. Any programme that cannot name, in advance, a specific decision it expects to change should not start. But the objection assumes the intuition is shared and stable, and it usually is not. Ask four experienced executives, separately, to assess the same live case and the spread is routinely far wider than any of them expects. Instrumentation does not replace experienced judgement. It tells you how much your experienced judgements differ from each other, which is a question intuition is structurally unable to answer about itself. And executive intuition covers the decisions executives see; the great majority are taken several levels below, and increasingly by software, at a volume no one is reading.

Risk	Control in Red Bear: Enterprise Decision Intelligence
A correlation sold as a cause	Deviation is reported with the measure's own volatility and never as what the decision caused; confounders at equal weight; no figure below the attribution threshold (FR-OUT-10)
A prediction sold as an explanation	Forecasts carry confidence and precedent count and decline below the configured count; collapsed by default; the product's own models are governed for drift and bias per version (FR-SUP-03, FR-DEL-12)
Selection on the dependent variable	Outcomes attached at the class's interval for every decision, not when someone remembers (FR-OUT-01)
Reasoning contaminated by hindsight	Reasoning and confidence immutable once written; harvested records marked and excluded from quality scoring (FR-REC-12, FR-CAP-09)
Coverage that self-selects	Capture where decisions already happen; the portal is a route with the same contract, not a privileged path (FR-CAP-13)
Impact analysis as surveillance	Slices by region and type, never by name; process quality computed without access to outcomes; System Admin excluded from the explorer (FR-OUT-08, FR-OUT-05, FR-OUT-10)
An invented return figure	No figure sits between baseline and indicators; every effect reconciles to the register (FR-OUT-07, FR-OUT-03)

Sources

Every claim attributed to a source names it in the text as *author*, *work*, *section*; the full references are below. Statements without a citation are Red Bear's own position, developed for Red Bear: Enterprise Decision Intelligence. Quotations are limited to fifteen words.

Cited as	Source	Authority
Duke	A. Duke, *Thinking in Bets*, Portfolio (2018): resulting	Practitioner reference
Kahneman, Sibony and Sunstein	D. Kahneman, O. Sibony and C. Sunstein, *Noise: A Flaw in Human Judgment*, Little, Brown (2021): noise as unwanted variability in interchangeable judgements; the noise audit	Academic reference
Lovallo and Sibony	D. Lovallo and O. Sibony, *The case for behavioral strategy*, McKinsey Quarterly (March 2010): the 1,048-decision study and its design limits	Practitioner research
Blenko, Mankins and Rogers	M. Blenko, M. Mankins and P. Rogers, *Decide & Deliver*, Harvard Business Review Press (2010): quality, speed, yield and effort	Practitioner research
Book 1	D. Oliver, *Empower, Trust, Verify*, Book 1 (2026), chapters 2 and 9: the quality quadrant; Cartesian analysis	Red Bear
Red Bear	EDI-FSD-01 v3.0 draft; Red Bear Platform Functional Specification RBP-FSD-01	Red Bear

Glossary

Term	Meaning in this paper
Decision impact analysis	Measuring the actual effect of decisions on cost, revenue, risk and time, class by class, with every effect reconciled to the register
Description	How a class is actually decided; needs classification and coverage, no outcomes
Pattern recognition	Which recorded features move together; says nothing about direction
Prediction	The likely outcome given what is recorded at the moment of decision; no account of why
Causal inference	Whether deciding differently would have changed the result; needs planned variation
Confounding	The condition that produced the decision also affecting the result by another path
Survivorship	The records of decisions that went badly enough tending not to be there
Noise	Unwanted variability between judgements that were meant to be interchangeable; measured as the spread, not the average
Expected trajectory	What an object's measure would have done, computed from history before the decision only, stored with its method
Attribution confidence	How far a deviation can be linked to a decision; degrades with confounders, affected links, hierarchy distance and window length

Term	Meaning in this paper
Harvested record	A historic decision imported once, reconstructed rather than contemporaneous, marked and excluded from quality scoring

The series

This paper is one of ten. Each stands alone and each follows the same nine sections, so a reader who knows one knows where to look in the next.

Paper	Title	Theme
EDI-WP-01	Cartesian Decision Making	Learning what happened next
EDI-WP-02	The Cost of Organisational Indecision	The decision that never gets made
EDI-WP-03	Trust and Verify	Visibility, not control
EDI-WP-04	Delegation Without Losing Governance	The boundary and the room
EDI-WP-05	Building a Decision-Making Culture	Judge the decision, not the result
EDI-WP-06	Decision Impact Analysis	What a decision dataset can and cannot tell you (this paper)
EDI-WP-07	The Decision-Aware Enterprise	Decisions in flight
EDI-WP-08	Decision Making Anywhere	The channel is part of the governance
EDI-WP-09	When Micromanagement Is the Right Management	Oversight is a dial
EDI-WP-21	The Customer Journey	Connect, Observe, Govern, Learn, Compound

Red Bear Artificial Intelligence · Abu Dhabi, United Arab Emirates · info@redbearai.net · www.redbearai.net