



ICM
HOLDINGS INTERNATIONAL

RESEARCH & DECISION STUDIO

Decisions under uncertainty

Information is valuable only when it can change a decision enough to justify its cost.

Information value and decision science

METHODS / EVIDENCE / WORKED STUDIES

ENGLISH EDITION / 2026-10-10

THE DECISION

Decisions under uncertainty

An institution should commission evidence against a decision it can still make. This study develops an information charter, a transparent payoff perimeter, conditional signal reasoning and research purchase rules. Original illustrative studies show why a useful signal can be too expensive, why an informative signal can leave every action unchanged, and why a negative association can still support a sensible response. The same frozen fixtures run in the public Decision lab. The study makes risk-neutral expected monetary value explicit and explains where rights, implementation constraints, omitted states, delay, causal identification and risk preferences require broader analysis. It gives sponsors, committees and operators practical records for commissioning, stopping and challenging research without treating an attractive calculation as permission to act.

Questions this study answers

- Which decision, action owner and deadline make an evidence request actionable?
- What assumptions connect a signal to state probabilities, feasible actions and comparable payoffs?
- When does conditional action improvement justify the cost of a study?
- Which constraints, methodological limits and challenge records can overturn the numerical recommendation?

Scope and limits

Illustrative inputs explain mechanisms and are not actual project data or empirical estimates.

A deterministic model cannot establish implementation feasibility, legal effect, lender acceptance or future outcomes.

Cited public guidance is used only within explicitly stated source scope; institutional frameworks require local adaptation.

CONTENTS

Reading the study

01 More information is not always a better decision	4	13 Multiple uncertainties create conditional dependence	65
02 Define actions before collecting evidence	9	14 A pilot needs an identification argument	70
03 A probability has a provenance	14	15 Expert elicitation needs discipline	75
04 Payoffs must share a decision basis.....	19	16 Stopping research can be rational.....	80
05 Expected value is a conditional calculation	24	17 A committee must own its loss function	85
06 Perfect information is an upper bound	29	18 Evidence quality must survive publication	90
07 A test changes beliefs through likelihoods.....	33	19 Where expected information value is insufficient	95
08 Sample information can be less valuable than precision suggests.....	38	20 The strongest argument for broad exploratory research	100
09 The decision boundary matters more than average confidence	43	21 Appendix: probability and value dictionary ...	105
10 Delay is an information cost	48	22 Appendix: illustrative information studies	110
11 Information can fail through an implementation gap	54	23 Appendix: research commission worksheets..	118
12 Risk aversion changes the decision criterion	59	24 Appendix: evidence scope and an admissible-response packet	126

DECISION ARCHITECTURE

01

More information is not always a better decision

An information decision charter

Start with an action that can still change

A research request becomes a decision problem when an institution can name an action that could change after the result arrives. That action may be a commitment, a staged trial, a withdrawal, a redesign or a deliberate decision to wait. A report title is not an action. Neither is a general instruction to understand a market better. Before discussing a study design, the commissioning team should state what the result could permit, prevent or alter. This gives researchers a target and gives the committee a basis for deciding whether the proposed evidence is worth buying.

Consider a committee facing an uncertain opportunity. It can commit at scale, preserve a smaller position or hold its resources. The uncertain state affects the payoffs of each action. A proposed signal can change the committee's view of that state. The essential comparison is between the best feasible action with current information and the conditional actions that would be selected after each possible signal. The signal creates decision value through those action changes. A more detailed description of the uncertainty is useful only to the extent that it improves a relevant choice or serves another explicitly declared institutional purpose.

The practical difficulty is that organisations often commission a study before agreeing on the menu of actions. A report then arrives to an audience that cannot change the design, lacks the authority to change the contract, or has already spent the budget. The information may be intellectually useful while its commissioning logic has failed. A decision charter should therefore describe the action owner as well as the research question. A committee that can recommend a change but cannot authorise it must identify the subsequent approval path and test whether that path can complete before the opportunity expires.

An information charter also needs a counterfactual. The relevant alternative to commissioning a study is not ignorance in the abstract. It is the action the institution would take using the information it already has, including its existing uncertainty, operational safeguards and available options. If the team would already stage the investment, a study must

improve upon staging. If a binding condition already requires a safety investigation, a voluntary purchase calculation cannot cancel that requirement. If the team would hold, the study must explain why its possible findings could make a feasible alternative preferable.

This makes the timing of the charter consequential. Writing it after the results are known invites a retrospective justification of almost any study. An unfavourable result can be called a useful warning; a favourable result can be called confirmation; an ambiguous result can be called a reason for another report. Each interpretation may be reasonable in a particular context, but without a recorded prior decision and a conditional response rule there is no stable way to distinguish a useful commission from a research process that continually changes its stated purpose.

The charter should preserve the distinction between information and intervention. A test may reveal that an operating design is weak without making that design easier to change. A survey may identify demand without creating a supply capability. A technical pilot may demonstrate a component while leaving its integration unresolved. The resulting decision value depends on the responses that the institution can implement, rather than the responses an analyst can imagine. This is why the paper's frozen action menus are explicit. Their admissibility is assumed for illustration and remains an independent implementation question.

The same discipline applies to an apparent decision to do nothing. Holding resources may preserve flexibility but can also sacrifice an expiring option, a learning opportunity or an obligation to maintain service. Its payoff belongs in the same perimeter as the active choices. A zero hold payoff is a convenient illustrative baseline, not a universal claim that inaction has no consequences. If holding changes obligations, relationships or future access, the payoff matrix must represent those consequences or explain why they are handled outside the monetary comparison.

An actionable research question is narrower than a comprehensive account of a business. The committee may need to know whether a specified operating condition is likely enough to justify a commitment. It may not need a general industry forecast, an exhaustive technology history or a complete explanation of every uncertainty. Research scope should follow the decision's sensitivity. A narrow study that resolves a binding choice can be more useful than a broad study that improves understanding of variables that do not alter the selected action. Scope discipline also makes the claim support more reviewable.

This does not imply that every worthwhile research activity must have an immediate monetary return. Institutions may commission foundational science, public evidence, capability development or long-term exploration. Those purposes should be named directly. A short-horizon expected monetary value calculation is poorly suited to an exploratory programme whose action set is expected to change. The error is to attach a numerical purchase rule to one purpose while relying on a different purpose to justify the expense. A mixed commission should separate its decision-support component from its broader public or exploratory objectives.

In the illustrative useful-signal case, the decision is deliberately simple. A binary state affects the payoffs of Commit, Stage and Hold. A binary signal has supplied conditional likelihoods and a stated purchase cost. The canonical engine chooses the action with the greatest expected payoff before information, calculates the posterior in each possible branch and chooses again. This stripped-down structure makes the mechanism visible. It does not estimate an actual opportunity's likelihood, validate a study instrument, or establish that a committee has the authority to implement the selected branch.

The value comparison should be recorded before the study is sold as persuasive evidence. A supplier can describe a highly accurate instrument while leaving the client's response rule unspecified. A sponsor can request a detailed forecast while assuming a project must proceed regardless. A committee can welcome a useful report and still have no intention of changing its position. These are commissioning problems that accuracy alone cannot repair. The charter should ask what adverse finding would stop or modify the commitment and what favourable finding would justify an action that is currently withheld.

The adverse finding deserves particular attention because it reveals whether the institution has made an evidence-contingent decision. If no possible result can change the commitment, the study is serving confirmation, communication or compliance purposes. Those purposes may still be legitimate, but their cost should not be justified by a claim of improved action selection. A charter should allow the author to state that the intended use is communication, provided the report does not then imply that a conditional decision rule was tested when none existed.

The favourable finding can be equally revealing. A positive result may be insufficient because the organisation still lacks financing, a permit, an operating capability or a contractual right. In that case the action menu should include the feasible next step, such as a bounded engineering investigation, rather than a full commitment. This preserves the difference between an attractive state of the world and an executable institutional

response. The paper later shows how removing an unavailable action can materially reduce information value even when the signal itself is unchanged.

Finally, the charter should nominate a person who can challenge the loss function. An analyst can faithfully calculate a model that excludes the cost borne by another department or treats a protected interest as a compensable inconvenience. The person approving the research must know whose payoffs are represented, who bears omitted consequences and who may reject the proposed trade-off. Information value is conditional on that perimeter. A technically correct result can still be an institutionally unacceptable basis for action.

Evidence and scope

Jackson and coauthors connect research value to expected loss under a stated decision, distinguishing perfect and sample information.

Evidence: A guide to value of information methods for prioritising research in health impact modelling

A completed illustrative information charter

The charter below defines a reviewable commission. Its owners are functional roles in a hypothetical institution, not actual people or clients.

A completed illustrative information charter		
Field	Illustrative entry	Challenge before commissioning
Decision	Choose Commit, Stage or Hold	Can each action still be authorised?
State	Specified opportunity succeeds or fails	Are material alternative states omitted?
Evidence	A single binary signal	What validates the conditional likelihoods?
Action owner	Committee with delegated scope	Who approves a response beyond that scope?
Deadline	Before commitment becomes irreversible	Will the result arrive and be usable in time?
Loss perimeter	Incremental payoff to the stated institution	Whose costs or rights are outside this perimeter?

Commission against a response rule

Record the current preferred action, the possible evidence outcomes and the feasible response to each outcome. Refuse a decision-value justification when no admissible result can change the response. If the commission instead serves public evidence, exploration or a mandatory investigation, name that purpose and assess it on an appropriate basis. Keep implementation approval distinct from the evidence purchase decision.

1. Identify the current action using current evidence.
2. Record feasible conditional responses before seeing the result.
3. Confirm authority and the latest usable delivery date.
4. State any broader research purpose separately.

Research can reveal an action nobody anticipated

A fixed action menu can understate the value of exploratory research that creates a new design or partnership. The answer is not to claim that the fixed-menu calculation measures that creative benefit. Commission an exploratory phase with a stated discovery purpose, then update the action set and the decision model when an actual new option becomes available. Preserve the initial purpose so that option discovery is not retrospectively invented to defend every study.

The charter establishes a perimeter, not the truth of its assumptions

Naming actions, states and owners makes a commission reviewable. It does not prove that the state model is complete, that payoffs are comparable, that a signal is valid or that implementation is feasible. Those are separate evidence obligations. The numerical studies later in this paper assume their inputs explicitly so that the reasoning can be examined without implying an empirical validation.

Decision implications

- An information request needs a current action and feasible conditional responses.
- A report can serve a worthwhile purpose without creating immediate decision value.
- Authority, timing and the payoff perimeter belong in the commissioning charter.

DECISION ARCHITECTURE

02

Define actions before collecting evidence

An action and outcome register

Define actions before evaluating evidence

An action menu is a statement about what the institution can do, not a list of outcomes it hopes to achieve. Commit, Stage and Hold are choices. Success and failure are uncertain states. Mixing the two produces a misleading model in which the committee appears to choose a favourable state merely by naming it. A reviewable decision table keeps actions in rows and states in columns. Each cell describes what follows if that action is taken and that state occurs, including the consequences that remain unpleasant even when the intended opportunity succeeds.

The menu should be assembled before asking an instrument to separate the states. Otherwise the signal may be designed to answer a question that no feasible action uses. For example, a technical assessment could discriminate between operating conditions while the only authorised response is an unchanged maintenance schedule. The study could still support a later redesign, but that later choice must be included with its actual decision owner and prerequisites. A proposed response should not enter the current model simply because it would be attractive if every constraint disappeared.

Actions should have comparable decision dates. Comparing an immediate full commitment with a staged action that includes several future discretionary choices can favour staging without accounting for the uncertainty in those future decisions. A simple model can still represent staging through a supplied aggregate payoff, provided the author states what has been folded into that number. If the staged strategy depends on future signals, changing costs or contingent rights, a richer sequential model is required to value it explicitly. The Decision lab does not calculate that sequence from a single Stage row.

The illustrative menu deliberately treats Stage as a complete supplied action with a payoff in each state. This permits a transparent comparison with Commit and Hold. It does not imply that an actual staged investment would have the same payoff or that its later phases would occur automatically. The model's action names are labels for the supplied

cells. Renaming Stage as Pilot would not turn its payoffs into a causal estimate of a pilot programme or add an option-pricing engine. The scope follows the computations and the documented assumptions.

A feasible action can be economically unattractive and still belong in the menu. Its presence allows the model to demonstrate why it is not selected or why a particular signal could make it preferable. Conversely, an economically attractive action that cannot be authorised should be excluded or marked as outside the current decision. Leaving it in the optimising menu and adding a footnote about feasibility does not correct the result. The numerical optimum would still depend on a response the institution cannot take. The restricted-action study makes this distinction concrete.

Removing Commit from the illustrative menu leaves the signal unchanged but changes the opportunities available after it. The positive branch can now lead only to Stage or Hold. This is a different decision problem, with a different perfect-information ceiling and a different research purchase result. The comparison is useful because it separates the quality of evidence from the value of the responses it unlocks. A strong research team cannot compensate for a client that has no executable action with which to exploit its findings.

The menu also needs an inclusion rule. A team can make its preferred proposal appear attractive by omitting an available alternative with lower downside. It can make research appear useful by excluding a robust action that would be chosen under either signal. It can make the current decision appear more uncertain by including fanciful options that have no sponsor or budget. An inclusion register should state who proposed each action, what evidence supports its feasibility and why plausible alternatives were excluded. This supports an adversarial review of the decision architecture before reviewers argue about probability.

The institution should distinguish a dominated action from an inadmissible action. An action is dominated within a stated payoff matrix when another available action is at least as good in every represented state and better in a relevant state. That is a property of the supplied comparison, not a factual claim about all real circumstances. An action is inadmissible because a rule, right, capability or mandate prevents it. A dominated action may be removed for clarity after checking the matrix; an inadmissible action must not be selected merely because its expected payoff is high.

Tie handling deserves a written policy. The canonical engine sorts exactly equal expected values by action name in stable literal code unit order used by JavaScript to produce deterministic output. This makes reproduction independent of browser language; it does not express institutional preference for the first displayed action. When a result lies at an intersection, the committee may choose on reversibility, operating burden or a protected constraint. Those reasons belong outside the numerical tie-break and, if they alter the decision, in a richer admissibility or preference analysis. A reproducible tie is not an ethical rule.

The action set should remain stable during a particular comparison. If a favourable signal prompts an author to add a new high-return action but an unfavourable signal uses the original restricted menu, the apparent information value combines learning with an undeclared change in opportunity. That may be a defensible model if the new action truly becomes available only after the signal, but its state and signal dependence need to be explicit. The simple app uses the same supplied actions in each branch. It does not create signal-contingent new options.

Reversibility is another property that cannot be read from an action name. A staged commitment may have an irreversible first tranche, a costly exit, reputational effects or an obligation to continue once an operating system is installed. Holding may preserve technical flexibility while losing a contractual opportunity. A committee should ask what can be reversed, by whom, at what cost and before which event. If these consequences are represented in supplied payoffs, their evidence belongs in the input register. If they are not, the limitation belongs in the decision condition rather than a silent narrative assumption.

The action owner should certify the response path, not merely approve the model's wording. For a proposed redesign, this means identifying the engineering authority, budget approval and interface responsibilities. For a withdrawal, it means identifying any contractual consequences and service obligations. For holding, it means confirming whether the institution may delay without breaching a duty or destroying the opportunity. This certification does not make the model true. It makes the meaning of its action labels sufficiently concrete for a committee to understand what it is actually being asked to approve.

The practical record is an action admissibility table that travels with the payoff matrix. It should distinguish currently executable actions from options that require another decision, and separate factual prerequisites from assumed future availability. An analyst can then run a provisional comparison with an explicitly broad menu and a second comparison with the currently authorised menu. The difference is an implementation sensitivity, not a measurement of the value of obtaining authority. The restricted study here illustrates that comparison without claiming to model an actual approval process.

There is a useful stopping rule at this stage. If the team cannot define a response to a plausible adverse result, it should repair the action architecture before buying precision. If it cannot identify an owner with authority to execute any conditional change, it should resolve the governance path or describe the study as preparation for a future decision. If it cannot represent the consequences of holding, it should challenge the baseline. These repairs can change the commission more fundamentally than refining a likelihood by another increment.

Evidence and scope

NASA frames risk-informed decisions through alternatives, performance measures and uncertainty within mission objectives.

Evidence: NASA Risk-Informed Decision Making Handbook

Illustrative calculation. Assumptions and sensitivity are published with the model.

Illustrative action menu and supplied payoffs

Every monetary value in this exhibit is a frozen input, a supplied scenario input.

Illustrative action menu and supplied payoffs			
Action	Success payoff, USDk	Failure payoff, USDk	Admissibility assumption
Commit	180.00	-100.00	Committee can authorise full commitment
Stage	70.00	-15.00	Aggregate staged payoff is supplied
Hold	0.00	0.00	Baseline opportunity costs are assumed represented

Illustrative calculation. Assumptions and sensitivity are published with the model.

Illustrative value with and without the unavailable action

The likelihoods are unchanged. Removing Commit changes the feasible response set.

Illustrative value with and without the unavailable action				
Menu	Best current value, USDk	Gross signal value, USDk	Net signal value, USDk	Decision reading
Commit, Stage, Hold	19.00	33.20	28.20	Purchase has positive model value
Stage, Hold	19.00	3.45	-1.55	Purchase exceeds model value

Freeze admissibility before optimising

Include an action only when the institution can describe its response path and the meaning of its supplied payoffs. Record an excluded alternative and its prerequisite rather than letting the optimiser select it with an informal feasibility caveat. Review a changed action set as a changed decision problem. Keep a deterministic computational tie-break separate from the committee reasons that resolve an institutional tie.

1. Separate choices from uncertain states.
2. Record an owner and prerequisites for each action.
3. Challenge excluded robust alternatives.
4. Recompute when the admissible menu changes.

A provisional menu can accelerate early exploration

An early team may not yet know which responses are executable. A provisional comparison can help identify valuable options to investigate. It should publish both its broad assumed menu and the currently available menu, clearly label the former as exploratory, and avoid treating its high expected value as an authorised commitment. The purpose is to find a governance or design question, not to conceal the unresolved action path.

Aggregate staging hides a sequence

The supplied Stage row compresses all later events into state payoffs. The engine does not calculate contingent phases, additional tests, future option exercise or approval delay. A real staged programme requires a dated decision tree or another appropriate sequential model when those mechanisms determine value. The simple comparison can expose a question for that analysis but cannot answer it by changing the row label.

Decision implications

- A feasible response menu determines the information that can create value.
- Removing an unavailable action changes the decision and may reverse the purchase result.
- The app tie-break ensures reproducibility and does not prescribe committee preferences.

DECISION ARCHITECTURE

03

A probability has a provenance

A probability evidence worksheet

A percentage needs an event and an evidence history

A probability input is a statement about an event, a population, a time horizon and an information set. Without those elements, a number can travel between decisions while changing its meaning. The prior probability in the illustrative engine refers to success under the stated binary opportunity definition before the proposed signal is observed. It is not a forecast accuracy score, a preference weight, an instrument sensitivity or the confidence an expert feels about a report. The input register should preserve this distinction so that a familiar percentage does not substitute for a specified event.

Success must be defined independently of the action labels. If Commit and Stage create different conditions for success, a common state may be inappropriate. The simple decision table assumes the state is shared across actions and only the payoffs differ. An actual intervention that changes the likelihood of success requires action-dependent state probabilities or a causal model. Supplying a larger success payoff for an active action does not make the engine account for its effect on the event probability. Reviewers should ask whether the proposed state is a fact about the opportunity or a result of the chosen intervention.

The probability's horizon should align with the payoffs. A chance of reaching an early technical milestone is not automatically the chance of delivering a lifetime commercial return. A long-duration programme may pass commissioning and fail later through operating, demand or regulatory mechanisms. A binary aggregate can be useful for a bounded illustration, but it compresses those mechanisms into a deliberately simplified state. If a committee needs to distinguish them, the event model must expand. The source of a prior for an early milestone cannot silently certify a broader success state.

The register should state how the number was produced. It may be a frequency from a defined sample, a model output, a structured expert judgement, a transferred estimate or an explicit scenario assumption. These origins have different review obligations. A frequency needs its

inclusion rule and denominator. A model output needs its inputs, calibration and limitations. An expert judgement needs its question, elicitation conditions and range. A scenario assumption needs its purpose and sensitivity. Calling every origin data prevents the committee from understanding which uncertainty has been measured and which has been supplied for analysis.

The frozen useful-signal study supplies its prior as an illustrative assumption. It is chosen to make the conditional decision mechanism visible. The paper does not infer that prior from a collection of client projects, claim that it is representative of an industry or estimate an actual commercial success rate. The purpose is methodological: a reader can change the input and inspect the resulting actions. The distinction matters because a perfectly reproducible calculation can only be as empirically applicable as the event definition and evidence supporting its inputs.

An estimate transferred from another context requires a bridge. A study of a mature operating system might not inform a first deployment; a likelihood measured with skilled operators might not apply to a different implementation team. A comparable label is insufficient. The bridge should explain the relevant differences in population, conditions, incentives and measurement. When those differences cannot be resolved, the author should report a range or retain the input as an assumption. The model can show whether the resulting uncertainty changes the decision; it cannot validate the transfer by arithmetic.

Probability provenance should not be confused with precision. Reporting more decimal places does not add evidence. A broad but honestly elicited range may be more decision-useful than a narrow point estimate with no traceable origin. The institution should retain enough precision to reproduce the calculation while displaying only the precision justified for interpretation. In an illustrative calculation, exact fixture values serve reproducibility, whereas the number of displayed decimals is a presentation convention. An actual evidence register would need a reason for the reported precision and an account of uncertainty in the estimate itself.

The prior also depends on what is already known. If the team has observed a precursor that is correlated with the proposed test, using an unconditional historical frequency as the current prior may count evidence incorrectly. A current prior should represent the institution's information before the new signal. The app does not reconstruct that information set or detect dependence with an earlier test. The author

must specify it. This is particularly important when a signal supplier reports an impressive historical accuracy that was evaluated after filtering opportunities using information already known to the client.

An expert's confidence in a probability is a different question from the probability itself. A panel can agree that an opportunity has a moderately favourable chance while having little evidence about that chance. It can also disagree despite substantial evidence because members use different event definitions. The register should separate the event estimate from the evidence assessment and disagreement record. This enables a committee to ask whether a point estimate is stable enough for the decision without treating a confidence adjective as a numerical likelihood. The paper uses no invented conversion between those concepts.

Measurement uncertainty introduces another distinction. An instrument may estimate a physical quantity with a stated uncertainty, while the decision requires the probability that a performance condition is satisfied. Turning the former into the latter requires a model of the quantity and the threshold, plus relevant uncertainty about measurement and context. The simple binary signal fixture supplies its conditional likelihoods directly. It does not derive them from a metrological uncertainty budget. A report that uses physical measurement precision to justify a commercial likelihood must make that bridge explicit and reviewable.

The register should identify the evidence that would change the prior before commissioning the proposed signal. This guards against a number becoming a fixed institutional belief. If new information contradicts its basis, the team should update the current decision and reassess the research purchase. A previously useful test may become redundant after another credible result arrives. Conversely, a new concern may increase the value of resolving the state. The charter is therefore a dated record, not a permanent certificate of how much a study is worth.

Sensitivity should concentrate on the decision boundary rather than only the input's plausible range. A committee may tolerate substantial uncertainty when the preferred action remains stable across that range. It may need better evidence when a small change selects a different action with consequential payoffs. The app's prior sweep gives a deterministic view of the best available action at sampled probabilities, and its pairwise intersections provide additional checks. These outputs support a decision sensitivity discussion. They do not supply the probability distribution of the prior or the evidential plausibility of each point.

Illustrative calculation. Assumptions and sensitivity are published with the model.

Completed illustrative probability provenance register

The record deliberately exposes the absence of empirical calibration.

Completed illustrative probability provenance register		
Field	Useful-signal fixture entry	Review obligation
Event	Defined opportunity success	Challenge whether binary states are sufficient
Prior	40.00 percent	Scenario assumption only
Population	No empirical universe declared	Do not infer an industry frequency
Information set	Before the supplied binary signal	Do not double count existing evidence
Horizon	Same aggregate horizon as payoff matrix	No early milestone to lifetime transfer
Owner	Commissioning analyst and action committee	Approve interpretation and sensitivity

The input owner should sign the probability's interpretation, not merely the numerical field. A record stating that the prior refers to the defined success state, the named opportunity and the pre-signal information set is more useful than a signature beside a percentage. The reviewer should then distinguish disputed interpretation from disputed magnitude. If participants disagree about the event, averaging their numbers can produce a precisely formatted estimate of no shared proposition. Resolving the definition should precede numerical aggregation.

There is also a legitimate unresolved state. An institution may lack a defensible probability model for a novel threat, an unprecedented interface or an ill-defined outcome. Entering a convenient middle value does not solve this problem. The committee can use scenario comparisons, robust constraints or a bounded exploratory commission while documenting the inability to support a probability. If it uses an illustrative prior to understand a mechanism, it should preserve that label. A decision-support app should help expose the missing evidence rather than pressure the user to disguise it as a forecast.

Evidence and scope

IPCC guidance separates confidence based on evidence and agreement from quantified likelihood in climate assessments.

Evidence: Guidance Note for Lead Authors of the IPCC Fifth Assessment Report on Consistent Treatment of Uncertainties

Evidence and scope

NIST distinguishes statistical and other evaluation methods within a measurement uncertainty account.

Evidence: Essentials of expressing measurement uncertainty: basic definitions

Accept a probability with its meaning

Require an event definition, horizon, information set, provenance and named owner before accepting a probability as an empirical input. Keep a transferred estimate provisional until its bridge is reviewed. When only a scenario assumption is defensible, label it and evaluate how the decision changes across alternatives. Reject invented precision and conversions from confidence adjectives into unsupported likelihoods.

1. State the event and horizon.
2. Record the pre-signal information set.
3. Identify frequency, model, judgement or assumption provenance.
4. Challenge transfer and boundary sensitivity.

A point assumption is useful early in a project

A team can learn from a provisional point input before collecting sufficient evidence. The important condition is that the provisional status travels with every interpretation. Use the point to expose the action mechanism, pair it with explicit alternative fixtures and record the evidence needed to replace it. A provisional number should not acquire empirical authority simply because later reports reproduce it accurately.

The engine does not infer input uncertainty

The app treats the supplied prior and likelihoods as fixed numbers. It does not estimate their confidence intervals, elicit expert distributions, infer action-dependent probabilities or validate a transferred measurement model. The probability register and qualitative uncertainty controls identify the analysis that would be required; they are not additional computations hidden behind the binary engine.

Decision implications

- Prior, likelihood, confidence and preference weight answer different questions.
- A scenario input can explain a mechanism without estimating an empirical success rate.
- Event definition and information provenance should be settled before numerical aggregation.

DECISION ARCHITECTURE

04

Payoffs must share a decision basis

A payoff matrix and basis test

Make the consequence account comparable

A payoff matrix is an account of consequences within an agreed perimeter. It is not merely a convenient set of numbers for an expected-value calculation. Every cell should specify whose consequences are represented, when they occur, what baseline they use and which material effects are excluded. The engine can compare supplied monetary cells reliably without knowing any of these facts. The committee must therefore review the perimeter before trusting the ranking. A correct multiplication cannot correct a payoff that omits the person bearing the downside.

Incremental payoffs require a common baseline. If Commit includes the cost of the investigation but Stage omits it, the comparison confounds the action with an inconsistent accounting boundary. If a historical sunk cost is charged to one action but ignored for another, the table can reward a choice for an expense that no present decision can avoid. The useful-signal fixture supplies state payoffs before the proposed test cost and treats that cost separately. This is an explicit modelling convention, not a claim about an actual accounting system or a universal rule for every research programme.

The currency unit also needs interpretation. The paper labels the illustrative payoffs USDk because the canonical input schema uses that unit. It does not imply that the opportunity is located in a dollar economy or that exchange-rate effects have been modelled. A real comparison involving several currencies would need a conversion basis, timing and scenario treatment. A label alone cannot make values commensurate. The same applies to nominal versus real amounts, present values versus undiscounted totals, and payments to the institution versus economic impacts on a broader public perimeter.

The payoff horizon determines what the state comparison can say. An aggregate lifetime present value and a near-term cash inflow are not interchangeable. A committee may care about both, but it should use distinct performance measures or a properly integrated model. The simple

app accepts aggregate supplied payoffs and does not discount a dated cash-flow schedule. Its output is an expected monetary comparison of those inputs. If delayed cash, financing pressure or recurrent operating effects dominate the choice, the author should calculate them in an appropriate canonical model or identify the missing analysis.

State payoffs should include the consequences of failure that the institution can actually bear. A failure cell might include an abandoned investment, remediation, service interruption or an exit cost. It might exclude a protected right that cannot lawfully be traded for a larger expected return. These exclusions have different meanings. A cash consequence excluded for lack of evidence creates an uncertainty gap; a non-compensable constraint creates an admissibility boundary. The committee should not resolve the latter by inserting an arbitrary monetary penalty and allowing the optimiser to trade it away.

Success can carry costs too. An action may succeed commercially while imposing operating burden, increasing concentration, creating a liability or consuming capacity needed elsewhere. Calling a cell success should not make it a gross revenue figure. The payoff should represent the relevant net consequence under that state. Where costs remain contingent within the success state, an aggregate may hide variation that matters to risk tolerance. The analyst should disclose that compression and test whether a richer state representation is required. A positive expected payoff does not establish tolerable downside exposure.

The hold baseline deserves the same diligence as an active option. The fixture supplies no monetary consequence for Hold in either represented state. This is a transparent illustrative simplification. In an actual decision, holding may preserve capital while losing a licence, a customer, a delivery slot or a learning opportunity. It may also avoid a cost that is unavoidable under commitment. If those effects are material, a zero baseline would misstate the action comparison. The input owner should explain the baseline rather than treating inaction as an inherently neutral alternative.

The proposed study cost is separated from the action payoffs in the app. This permits the model to display gross information value and then subtract the purchase price once. It also exposes a frequent perimeter error: a team may include the same research cost within an action payoff and subtract it again as the test cost. The paper's fixture convention avoids that double charge by stating that the action payoffs precede the proposed signal purchase. A real commission must review its cost categories so that acquisition, analysis and implementation charges are each included exactly where they belong.

Research costs can have several incidences. A vendor price is only one component. Internal analysts may need time, an operating trial may interrupt service, and a decision may wait while a report is reviewed. The binary engine has a single testCost field. It can accept an aggregate cost if the components have been calculated and documented elsewhere, but it does not estimate those components. A decision package should show how the aggregate was assembled, who bears it, and whether any important nonmonetary consequence is handled as a condition outside the expected-value comparison.

Payoffs are often the more contentious inputs because they express institutional priorities. Different parties may agree on state probabilities and still rank the actions differently when their consequences differ. A sponsor may bear construction cost while an operator bears downtime; a public institution may value continuity while a private party values cash return. The paper does not resolve those differences through a single universal monetary objective. It recommends a declared decision owner and payoff perimeter, with distributional consequences and protected constraints reviewed explicitly. The same evidence can have different value to different legitimate decision makers.

The original completed matrix below gives a functional account of each cell. It does not invent contracts, clients or actual estimated returns. Its purpose is to ask how the supplied input would be supported in a real dossier. For Commit, the review would need the incremental success case and failure exposure. For Stage, it would need an account of the compressed subsequent decisions. For Hold, it would need the opportunity cost and any obligation that persists. For the study cost, it would need a distinct commissioning budget and cost boundary.

Sensitivity should test coherent alternative payoffs, rather than changing isolated numbers without considering their meaning. A reduced upside could reflect a smaller opportunity, a delay already embedded in supplied aggregate values, or a more conservative commercial case. Each interpretation changes the question. The paper's reduced-upside fixture simply supplies a different matrix and discloses that convention. It does not calculate a time discount or infer how long a study takes. Coherent inputs are more important than the attractive appearance of a sensitivity chart.

Completed illustrative payoff perimeter

This is an input acceptance record, not an empirical valuation.

Completed illustrative payoff perimeter			
Input	Included by assumption	Excluded or unresolved	Evidence needed in a real dossier
Commit success	Incremental net monetary consequence	Distribution and implementation details	Coherent commercial and operating case
Commit failure	Supplied downside monetary exposure	Actual legal remedies and timing	Failure mechanism and liability review
Stage cells	Aggregate staged strategy payoff	Explicit future option sequence	Phase decision tree and feasibility
Hold cells	Illustrative baseline	Actual opportunity costs	Rights, deadlines and counterfactual
Study cost	One aggregate purchase cost	Separate delay engine	Cost component register
Protected interests	Admissibility review outside payoff	Arbitrary monetary substitution	Appropriate authority and evidence

The committee should also record why it accepts a monetary objective for this particular choice. A risk-neutral expected-value rule can be useful for repeated, diversifiable or limited exposures, but the app cannot prove that these conditions hold. A single concentrated decision can pose liquidity or mandate constraints that an aggregate payoff does not capture. The appropriate response may be a constraint, a utility analysis, a stress test or a different action set. These additions should be specified as additional methods, rather than assumed to emerge from a high expected value.

An acceptable payoff matrix therefore has both a numerical and a narrative audit trail. Numerical checks establish units, arithmetic and consistent cost placement. The narrative trail establishes incidence, timing, baseline, exclusions and reasons for the chosen objective. A reviewer should be able to disagree with a cell while understanding exactly what it means. That is a stronger foundation for a committee debate than an apparently precise matrix whose numbers were agreed only because nobody asked what they excluded.

Evidence and scope

HM Treasury separates objectives, options, costs, benefits and uncertainty in public appraisal.

Evidence: [The Green Book](#)

Evidence and scope

UK multicriteria guidance addresses preference judgements when several decision criteria require comparison.

Evidence: Multi-criteria analysis: a manual

Illustrative calculation. Assumptions and sensitivity are published with the model.

Illustrative cost perimeter and value comparison

The useful and expensive signals use the same payoffs and likelihoods. Only the supplied purchase cost changes.

Illustrative cost perimeter and value comparison			
Fixture	Cost, USDk	Gross signal value, USDk	Net value, USDk
Useful signal	5.00	33.20	28.20
Expensive signal	80.00	33.20	-46.80

Approve the loss perimeter before the result

Record units, baseline, timing, incidence and material exclusions for each payoff. Place the proposed study cost once. Separate missing monetary consequences from protected conditions that determine admissibility. Recompute with coherent alternative matrices when the opportunity changes, and require another method for consequences the supplied aggregate cannot represent.

- 1. Trace each payoff to its consequence perimeter.
- 2. Challenge the hold baseline and failure exposure.
- 3. Separate test cost from pre-test action payoffs.
- 4. Escalate non-compensable constraints.

A rough payoff matrix can still reveal a robust choice

Early estimates can support a useful screening comparison if the decision remains stable under coherent challenges to the values. This is a reason to test the perimeter and range, not a reason to present rough cells as settled evidence. Record the conditions under which the screening result would fail and the consequence evidence that must be obtained before an irreversible commitment.

The app compares aggregate monetary inputs

The engine does not discount dated cash flows, resolve distributional conflicts, calculate liquidity constraints or monetise protected rights. The unit label and expected-value arithmetic do not establish those properties. The comparison is conditional on the supplied matrix being suitable for the stated institutional decision; unresolved perimeter questions can invalidate its recommendation even when every output is reproduced.

Decision implications

- A monetary cell needs a common baseline, horizon, incidence and cost boundary.
- Study price should be charged once and distinguished from the action payoffs.
- Protected constraints determine admissibility and require the appropriate authority.

MECHANISMS AND WORKED STUDIES

05 Expected value is a conditional calculation

A canonical decision app baseline

Choose a benchmark before valuing a signal

Expected monetary value is a conditional comparison of supplied consequences. For each action, the engine weights its success payoff by the prior success probability and its failure payoff by the complementary probability. It then chooses the largest expected payoff among the supplied actions. This is a transparent risk-neutral rule. Its clarity is valuable because a reviewer can identify the exact assumptions that determine the ranking. Its limitations are equally clear: a high average does not by itself establish acceptable risk, available cash or permission to take the action.

The useful-signal fixture illustrates why the most attractive success payoff need not define the current choice. Commit offers greater upside in the success state but also greater exposure in failure. Stage has a smaller upside and a smaller downside. Hold has the supplied baseline payoff. At the stated prior, the engine selects Stage before receiving new evidence. The current action is therefore a compromise within the supplied monetary objective, not an informal preference for caution. A different prior or payoff matrix could make another action best, as the later sensitivity outputs show.

The current expected payoff is the benchmark for information value. The model does not compare the signal-assisted policy with a randomly selected action or with the action the study supplier hopes to recommend. It compares it with the best action under current information. This benchmark matters because research can appear very valuable when

evaluated against an unnecessarily poor baseline. A committee should first repair the current choice using the evidence it already has. Only then should it pay for information that improves upon that choice.

An expected payoff can be understood as an average over the represented uncertain states under the supplied probabilities. It is not a promise that the institution will receive that amount in a single realised decision. In the fixture, the selected Stage action will receive its success cell or its failure cell if the binary model describes the outcome. The expectation is a planning comparison, not another realised state between them. A decision package should preserve the state payoffs alongside the average so that the downside does not disappear behind a single metric.

The rule is linear in each action's probability-weighted payoffs. This makes it responsive to the prior but does not make it empirically calibrated. A probability assumption with no basis remains an assumption after multiplication. A payoff estimate with an omitted cost remains incomplete after averaging. The expected-value exhibit is therefore a good place to challenge the input perimeter. Reviewers can ask which cell or probability would need to change for the preferred action to change, then direct evidence work to those consequential uncertainties rather than indiscriminately refining every input.

The ranking should be reported with the available alternatives. Publishing only the winning metric hides how close the next action is and what trade-off produced the result. A narrow difference may depend on an input that is poorly supported; a larger difference may survive a substantial challenge. The paper's tables retain the action payoffs and expected values. Later policy and intersection outputs make the sensitivity more explicit. The app does not attach a statistical significance test to a difference in illustrative expected values, and the report should not invent one.

Risk neutrality is a decision assumption, not a claim that the institution feels no risk. The rule values monetary outcomes linearly within the supplied perimeter. An institution with concentrated exposure, binding solvency requirements or a different utility objective may prefer an action with lower expected money. Those preferences need to be represented through an admissibility constraint or another declared method. The expected-value model can still provide a useful reference comparison, but its optimum should not be presented as the only rational decision regardless of mandate and exposure.

The choice of unit does not alter the ranking when all payoffs are scaled consistently by the same positive factor. This is a property of the supplied linear comparison. Changing only a subset of costs or mixing units

changes the decision problem instead. A practical review should therefore check dimensional consistency before debating the result. It should also distinguish conversion from valuation: converting an amount to a common unit does not prove that different consequences have been assigned acceptable monetary values. The committee's perimeter approval remains necessary.

The state probabilities should exhaust the represented universe. In the binary fixture, success and failure are complementary by construction. If a real opportunity includes abandonment, partial delivery, regulatory suspension or a mixed outcome that differs materially from either cell, those consequences have been compressed or omitted. A binary model may still be a useful screen when the compression is justified. It should not be used to hide an outcome that would overturn the decision. The source of each aggregate cell should identify what is included in its state and what remains outside it.

The app also reports expected regret. Regret compares the payoff from an action with the state-specific best payoff that would have been available if the state were known. Its expected value uses the same supplied probability model. This gives another interpretation of the current choice: choosing the largest expected payoff minimises expected regret within this complete fixed menu. It does not mean the institution will avoid disappointment, and it does not establish a maximum-regret decision rule. Those are different interpretations that must not be inferred from the label.

The current optimum is dated. If the opportunity changes while a study is commissioned, the original expected-value benchmark can become irrelevant. The reduced-upside fixture uses new supplied payoffs to illustrate that point. It does not claim to calculate how much value time has destroyed. Instead it shows the need to rerun the comparison whenever the input perimeter changes. A committee should preserve the original fixture and the reason for a revision so that a new output does not erase the basis on which the study was initially approved.

Arithmetic equality and reporting precision need separate rules. The current engine constructs reduced rational values from the canonical decimal forms of JSON-number inputs. It compares action totals, information values and net purchase sign before converting results to numeric display. Algebraic cancellation therefore gives exact zero for an unchanged-policy information value; the complete JSON retains the exact numerator and denominator as well as the display value. Display rounding can still make a small positive or negative amount look like zero. Reviewers should inspect the exact sign before discussing a substantive

decision margin. Equal action totals receive a stable literal text order for presentation, which gives no economic priority to the first displayed name. Exact arithmetic does not validate the supplied probabilities, payoffs or authority.

Expected regret provides a useful cross-check on the current action comparison. With the same prior, menu and payoff matrix, the state-specific best-payoff benchmark is common to every action. Subtracting each action's expected payoff from that common benchmark reverses the ordering: the highest expected payoff has the lowest expected regret. If a published table appears to disagree, first inspect whether it uses the same menu, units and state definition, or whether rounding has obscured equality. This check tests interpretation and object identity without introducing a second decision objective. A committee can still reject the monetary perimeter, but it should distinguish that judgement from a disagreement between two displays of the same supplied account.

The cross-check also identifies what information can change. A signal can reduce expected regret by enabling a different action in different states; it cannot eliminate a consequence excluded from the supplied matrix. If a hypothetical committee adds a protected-service condition, the feasible menu may change before any probability changes. Recompute that menu rather than adjust the prior until the expected-value leader matches the desired outcome. If the objection concerns missing consequences, repair the account instead. These branches locate a challenge in probability, consequence or admissibility and keep the reproduced monetary benchmark available while the institution resolves the disputed premise.

Evidence and scope

The Aqua Book requires proportionate assurance and explicit treatment of analytical limitations.

Evidence: The AQUA Book

Expected payoff before information

Expected payoff for action a = $p \times \text{success payoff for } a + (\text{one minus } p) \times \text{failure payoff for } a$. The current decision value is the greatest expected payoff among the admissible supplied actions. The formula uses a shared binary state and a risk-neutral monetary objective.

Illustrative calculation. Assumptions and sensitivity are published with the model.

Illustrative current choice, with its state consequences

The current ranking is computed using the useful-signal prior. State payoffs remain visible beside their expectations.

Illustrative current choice, with its state consequences			
Action	Success, USDk	Failure, USDk	Expected payoff, USDk
Stage	70.00	-15.00	19.00
Commit	180.00	-100.00	12.00
Hold	0.00	0.00	0.00

Accept the optimum conditionally

Review the current best action before assessing new information. Retain alternative values and state consequences, then challenge the probability, perimeter and objective where a change could alter the choice. When risk preferences, protected constraints or cash requirements conflict with the monetary optimum, record the conflict and use the appropriate additional analysis rather than treating expected value as an automatic approval.

- 1. Check complementary states and common units.
- 2. Publish alternatives beside the current optimum.
- 3. Challenge decision-sensitive inputs.
- 4. Record constraints beyond the risk-neutral rule.

The best average can be unacceptable in a single exposure

A concentrated institution may be unable to survive an adverse state even when the expected payoff is attractive. That objection is compatible with the arithmetic. It identifies a missing admissibility or preference condition. Analyse the exposure with a suitable liquidity, stress or utility method, and prevent the optimiser from selecting an action that breaches a binding institutional constraint.

Expectation is not an outcome guarantee

The expected monetary comparison does not estimate the distribution of cash within each state, establish solvency, provide a utility function or validate the prior. It represents the binary supplied consequences and the chosen objective. A reported average should never replace the state exposures and evidence limitations in a decision package.

Decision implications

- The information benchmark is the best current action under the declared rule.
- Expected money remains conditional on the states, inputs, admissibility and risk-neutral objective.
- Retain raw outputs for replication and distinguish finite precision from a consequential tie.

MECHANISMS AND WORKED STUDIES

06

Perfect information is an upper bound

A value-of-perfect-information worked study

Know the ceiling before buying the instrument

Perfect information is a deliberately strong counterfactual. It imagines that the institution learns the represented state before choosing an action and can then select the best available payoff in that state. The comparison preserves the same action menu and state probabilities as the current decision. It changes only the information available at the moment of choice. The resulting gain over the best current expected payoff is the expected value of perfect information. It is a ceiling within the declared model, not a product that a research supplier can ordinarily deliver.

The ceiling is useful because it can rule out an expensive commission without requiring a detailed model of the proposed instrument. If even perfect state knowledge could not improve the action enough to cover the study's cost, an imperfect study about that same state cannot justify that cost under the same fixed action menu and monetary rule. This screening argument is conditional. It does not apply unchanged if the study creates a new action, serves a mandatory purpose or improves a broader decision beyond the one represented. The charter should establish that scope before using the ceiling.

In the useful-signal fixture, perfect knowledge would allow Commit when the opportunity succeeds and Hold when it fails. The current action is Stage because the state remains uncertain. Perfect information therefore avoids the downside associated with acting in the unfavourable state while preserving the largest success payoff. The engine weights those state-specific best payoffs by the prior and subtracts the current best expected payoff. It does not assume that the committee can control the state, and it does not award the success payoff in every outcome.

The ceiling should be reported with its construction. A large number can look like the value of commissioning a study when it actually describes a hypothetical instrument with complete state knowledge. The report should identify the current action, the state-specific best actions and the resulting information gain. The reader can then see what an imperfect signal would need to accomplish. A supplier that resolves only a technical

component does not deliver perfect information about an aggregate commercial state unless a defensible relationship connects that component to the state.

The same ceiling can be understood through expected regret. The perfect-information policy defines the best payoff available in each represented state. The current action sacrifices some payoff relative to that policy when the state turns out differently from what would favour its choice. The current optimum's expected regret equals the perfect-information gain in this fixed-menu model. This interpretation helps a committee ask where preventable decision loss arises. It does not imply that all disappointment is avoidable through research or that regret is an observable financial charge.

The restricted-action fixture makes the ceiling's dependence on feasibility visible. Without Commit, perfect information can choose Stage in success and Hold in failure. The upside available after learning is smaller. The instrument has the same conditional likelihoods as before, but the information cannot recover an action that the institution is not allowed to take. This reduces the ceiling and can change the purchase decision. A research value presentation should therefore avoid using the unrestricted ceiling when the committee's actual response menu is narrower.

The ceiling is also sensitive to the payoff perimeter. If a failure cell omits a material consequence, avoiding that state through informed action may appear less valuable than it should. If a success payoff includes an unexecutable opportunity, perfect knowledge may appear more valuable than it could be in practice. The bound inherits the input assumptions rather than correcting them. A screening decision should therefore review whether the state-specific maxima are legitimate available actions, not merely take the largest supplied cells without examining their meaning.

A small ceiling does not mean the uncertain state is easy to predict. It can mean that the actions have similar consequences or that one robust action already performs well across both states. Conversely, a large ceiling can coexist with little practical prospect of obtaining useful information. The state may be difficult to observe before commitment, the relevant evidence may arrive too late, or the available instrument may have weak discriminatory power. The ceiling identifies the maximum decision loss that information could remove in the model. It does not estimate what any particular study will remove.

This distinction can guide research priorities. When the ceiling is small relative to a plausible commission, the team can focus on acting well with existing information or redesigning the action menu. When it is

substantial, the team can investigate whether a feasible signal can capture enough of that value. The order is economical: establish the decision and information ceiling, then describe the proposed evidence mechanism and cost. It prevents a team from spending effort optimising an instrument for a decision whose response opportunities cannot support the expense.

The perfect-price-boundary fixture deliberately uses a fully informative signal and a purchase price equal to its gross decision gain. The net information value is at the model's purchase boundary. This does not mean the real world offers perfect observation at that price. It is a transparent illustrative case that tests the relationship between a ceiling and a price. At the boundary, an institution would need to examine omitted costs, timing, implementation and rounding before making a commitment. A calculation with no margin should not conceal unresolved practical expenses.

The upper-bound argument has a clear reason within the model. With perfect state knowledge, the policy can choose a payoff at least as large as any available action in that state. An imperfect signal partitions uncertainty but cannot create a payoff beyond those state-specific maxima. Averaging the conditional policies therefore cannot exceed the perfectly informed expected payoff when the state model, menu and payoffs remain fixed. The engine preserves that structure. The report should state the conditions rather than present the inequality as a bound on all kinds of research benefit.

Information can also have value across several decisions. A measurement may inform maintenance, procurement and a later expansion. A ceiling calculated for only the immediate commitment omits those other uses. A programme-level appraisal could model them if it specifies the decisions, information reuse, timing and overlapping benefits. Summing separate ceilings without checking their shared payoffs and constraints risks double counting. The simple binary study does not calculate a portfolio of research benefits. Its limited scope is a reason to define the commission carefully, not to dismiss every broader purpose.

Use the ceiling as a dominance screen on price, rather than a spending allowance. In a hypothetical commission priced above the same-menu perfect-information gain, even a flawless signal cannot recover its acquisition burden through the represented decision alone. Improving sensitivity or specificity cannot repair that price branch. A lower quote moves the enquiry to actual signal performance and response feasibility; it does not establish a purchase. A separate mandatory or public purpose can justify research on another basis, but that purpose must be stated

independently rather than credited to the ceiling. This disposition makes rejection above the bound useful without turning every below-bound proposal into an apparent bargain.

A stale ceiling can be diagnosed by the benchmark that changed. If a hypothetical institution loses access to its high-upside response, both the current optimum and state-specific optima must be recomputed from the remaining menu. If a new prior arrives from evidence already paid for, the ceiling must use that information set rather than the commissioning prior. If only the vendor price changes, the gross ceiling need not change, although the purchase conclusion may. Keeping these branches separate tells the reviewer which object needs reproduction. It prevents an outdated maximum from surviving simply because it was once correct, while avoiding unnecessary reconstruction of an unchanged payoff account.

Evidence and scope

Jackson and coauthors distinguish perfect-information value from the benefit of a proposed sample study.

Evidence: A guide to value of information methods for prioritising research in health impact modelling

Perfect-information ceiling

Perfect-information expected payoff = $p \times \text{greatest available success payoff} + (\text{one minus } p) \times \text{greatest available failure payoff}$. Perfect-information value = perfect-information expected payoff minus best current expected payoff. The same states, prior, action set and payoff perimeter must be retained.

Illustrative calculation. Assumptions and sensitivity are published with the model.

Illustrative information ceilings depend on the action set

The computation chooses the state-specific maximum within each supplied menu.

Illustrative information ceilings depend on the action set			
Fixture	Current value, USDk	Perfect value, USDk	Perfect-information gain, USDk
Useful signal	19.00	72.00	53.00
Restricted action	19.00	28.00	9.00
Perfect price boundary	0.00	9.00	9.00

Use the ceiling to screen, then model the actual instrument

Calculate a perfect-information bound only after accepting the current decision perimeter. A commission above that bound cannot be justified by improving this fixed decision under the stated objective. A commission below it still requires an instrument-specific calculation. Record broader legitimate uses separately and review overlapping benefits before adding them to an information budget.

1. Confirm state-specific actions are feasible.
2. Retain the same prior and payoff perimeter.
3. Reject prices above the applicable bound.
4. Evaluate the actual signal before approving a lower price.

A study may create new opportunities

Research that discovers a new action can exceed a ceiling constructed from the old menu because it changes the decision problem. That is a valid objection to applying the bound outside its scope. It is not a reason to call the fixed-menu ceiling wrong. Describe the exploratory purpose and the opportunity-creation mechanism, then reassess an actual expanded menu when the new action is sufficiently defined.

Perfect information is a hypothetical state-observation policy

The app computes a bound for the supplied binary state and fixed actions. It does not estimate whether complete state knowledge is attainable, how much a real instrument captures, or the benefits of research across multiple decisions. Those questions require their own evidence and model. The ceiling provides a disciplined screen within the declared perimeter.

Decision implications

- A perfect-information bound can reject an overly expensive commission within a fixed decision.
- A price below the ceiling still needs an instrument-specific value test.
- Action restrictions and payoff changes alter the ceiling.

MECHANISMS AND WORKED STUDIES

07 A test changes beliefs through likelihoods

An illustrative test evidence tree

Update from likelihoods, not signal labels

A signal is useful when its possible outcomes have different relationships to the state the institution needs to understand. The binary fixture supplies sensitivity as the probability of a positive signal given success and

specificity as the probability of a negative signal given failure. These are conditional likelihoods, not the probability that success follows a positive signal. The latter depends on the prior as well. Confusing these directions can make an instrument look more persuasive than the supplied assumptions justify.

The engine first calculates the probability mass of each signal branch. A positive signal can arise from success with a positive response or from failure with a false positive. A negative signal can arise from success with a false negative or from failure with a negative response. The branch mass adds the relevant prior-weighted possibilities. It then calculates the posterior success probability within the branch by dividing the success contribution by the total branch mass. This is a transparent conditional update under the supplied binary state and likelihood assumptions.

The useful-signal case makes the distinction visible. The positive branch raises the success probability enough to select Commit, while the negative branch lowers it enough to select Hold. The update does not produce certainty in either branch because the instrument is imperfect. The report should retain both posterior probabilities and branch masses. A highly favourable posterior in a rare branch may contribute less to expected decision value than a more moderate posterior in a common branch. Conditional appeal and ex ante value are different readings.

The false-positive-burden case changes the prior and weakens the negative-state filter. A positive signal can still be more likely when success occurs than when failure occurs, but failure is sufficiently common in the supplied state mixture that positive signals do not necessarily justify commitment. In this fixture, both signal branches select Hold. The instrument updates beliefs without creating a preferred action change. This is an example of why sensitivity alone cannot determine a commercial research purchase. The initial incidence and the response thresholds matter.

Signal names should not supply an interpretation that the likelihoods contradict. Positive is an instrument label, not a promise of favourable evidence. In the inverse-signal study, a positive result reduces the success probability while a negative result increases it. The supplied likelihoods make the signal negatively associated with success. The engine uses that association correctly: the decision response is attached to the posterior, not the emotional connotation of the word positive. An institution must train users to read the instrument's actual conditional meaning.

This inverse case is a useful adversarial test because it exposes a hidden assumption in many presentations. A decision tree can look convincing while its branches have been labelled in a way that presupposes the desired conclusion. If the analyst follows the word positive rather than the likelihoods, the same instrument can lead to the wrong action. The paper deliberately preserves the counterintuitive fixture and its raw outputs. Its purpose is to test method interpretation, not to suggest that a poorly designed or negatively associated instrument should be purchased without validation.

The likelihood inputs require their own evidence provenance. An instrument supplier may report accuracy over a population whose state mix differs from the intended use, or may estimate performance against an imperfect reference. The app does not inspect that study design. Its supplied sensitivity and specificity are assumptions unless an actual validation record supports them for the relevant state and use conditions. A commissioning dossier should separate the instrument's development evidence from evidence of transfer to the current opportunity and from the probability estimate used before the signal.

The signal should also be independent of the action in the manner the model assumes. If committing changes the measurement conditions or the state, a single likelihood table may not describe every action. If observing the signal requires an operating trial that itself changes the opportunity, information acquisition and intervention are intertwined. The binary engine treats the supplied signal as a source of evidence preceding the selected action. It does not identify a causal effect of testing or model a measurement process that changes its own state.

An impossible branch needs explicit treatment. If its probability mass is absent under the supplied assumptions, there is no posterior distribution conditional on that event in this model. The engine reports the posterior as unavailable and gives the branch no expected-value contribution. A reviewer should not replace that unavailable value with an arbitrary success probability to make a table complete. If the allegedly impossible outcome actually occurs, the event is evidence that the assumed state or likelihood model needs review. It is not a routine branch to which the same model can assign a confident response.

The branch probabilities should sum to the full represented signal universe. In the binary fixture, positive and negative are exhaustive outcomes. A real instrument may produce missing, ambiguous, delayed or invalid results. If those outcomes have different action consequences, collapsing them into positive or negative can distort value. The institution should either justify the collapse or use a broader signal model. The simple

app does not estimate the probability of an indeterminate result. Its illustrative likelihoods should not be used to hide an unresolved test-operating condition.

A posterior should remain associated with the specific fixture that produced it. Copying a favourable branch probability into another payoff matrix or opportunity without preserving the prior and likelihood context can create a misleading decision narrative. The study's downloadable inputs, outputs and bindings make that association concrete. A reader can reproduce the branch and see the supplied assumptions. If the context changes, the analyst should create a new fixture or revise the actual decision record with an explicit history, rather than reuse a posterior as a general property of the instrument.

The update should be checked in both directions. A team may inspect the favourable branch carefully because it supports the desired action while treating the adverse branch as a formality. Yet information value depends on the full conditional policy, including the action that avoids unnecessary exposure after an adverse result. A study that can only confirm a commitment and has no permissible adverse response may create less value than its presentation implies. The action charter and branch table should be reviewed together so that the institution's response is genuinely conditional.

A signal-label audit can distinguish a reporting error from an instrument error. In the inverse fictional case, the label positive is less associated with success than negative. Swapping display names while preserving the joint masses should swap the branch labels and leave the underlying conditional policy value unchanged. Swapping the likelihood values without the labels changes the instrument and requires a new comparison. If a provider's report uses favourable where the specification uses positive, reconcile the mapping before dispatching a response. The audit follows the conditional evidence object, not the emotional meaning of its name, and does not validate the likelihoods or their use population.

A result outside the binary specification needs a different disposition from an impossible model branch. A hypothetical inconclusive report can occur in practice even though the current instrument object has no inconclusive outcome. Do not force it into negative, renormalise the remaining results or fabricate a posterior. Retain the current fallback action while obtaining the interpretation or a model that represents the additional outcome. By contrast, a branch with zero supplied mass has an undefined posterior within the declared model; it contributes no weighted value but remains a

diagnostic of an incompatible observed result if it actually occurs. These distinctions protect the response rule at the handoff where a neat tree meets an imperfect report.

Evidence and scope

FDA diagnostic guidance defines sensitivity and specificity conditionally and requires interpretation in the intended population and study context.

Evidence: Statistical Guidance on Reporting Results from Studies Evaluating Diagnostic Tests

A conditional positive branch

Positive branch mass = $p \times \text{sensitivity} + (\text{one minus } p) \times (\text{one minus specificity})$. Posterior success after a positive signal = $p \times \text{sensitivity} / \text{positive branch mass}$, when that mass is positive. The negative branch uses the complementary conditional likelihoods. An absent branch has no defined posterior in this model.

Illustrative calculation. Assumptions and sensitivity are published with the model.

Illustrative useful-signal branch account

Probabilities are displayed as percentages; the frozen output retains fractions.

Illustrative useful-signal branch account					
Signal	Branch mass	Posterior success	Chosen action	Conditional payoff, USDk	Weighted payoff, USDk
Positive	43.00%	79.07%	Commit	121.40	52.20
Negative	57.00%	10.53%	Hold	0.00	0.00

Illustrative calculation. Assumptions and sensitivity are published with the model.

Illustrative inverse signal: names do not determine meaning

The supplied likelihoods reverse the intuitive favourable reading of a positive signal.

Illustrative inverse signal: names do not determine meaning			
Signal	Branch mass	Posterior success	Chosen action
Positive		50.00%	16.00% Hold
Negative		50.00%	64.00% Commit

Validate the conditional interpretation before using the branch

Require the intended state, likelihood direction and validation context to be explicit. Inspect both possible branches and their feasible responses. Treat an unavailable posterior as a model boundary, and review unrepresented outcomes rather than inventing their probabilities. Read a signal from its supplied association with the state, not from the connotation of its label.

1. Separate likelihood from posterior.
2. Inspect prior and reference-population compatibility.
3. Publish branch masses and conditional actions.
4. Escalate unrepresented or impossible outcomes.

A simple signal can communicate more clearly than a detailed instrument model

A binary abstraction can be appropriate when its decision consequences and validation bridge are defensible. Simplicity becomes misleading when indeterminate outcomes, context dependence or action effects are omitted despite changing the response. Preserve the simple model only after recording why the omitted mechanisms do not control the decision, or use it as a labelled screen pending a richer instrument account.

The instrument likelihoods are supplied

The app does not estimate sensitivity or specificity from study data, correct reference-standard bias, model indeterminate results or detect dependence with prior evidence. FDA material is used only for narrow diagnostic reporting definitions and context; it does not validate this commercial illustration or establish medical advice. The branch computations remain conditional on the declared likelihood assumptions.

Decision implications

- Likelihood and posterior run in different conditional directions.
- An instrument label does not guarantee a favourable update.
- Branch mass, posterior and feasible response belong in the same record.

MECHANISMS AND WORKED STUDIES

08

Sample information can be less valuable than precision suggests

A value-of-sample-information comparison

Separate usefulness from the purchase decision

The expected value of a proposed signal is the improvement in the expected payoff of the conditional policy it enables. The engine chooses a best action after each possible signal, weights that action's conditional

payoff by the branch probability and adds the contributions. It then subtracts the best current expected payoff. This gross value describes how much the signal improves this decision before acquisition cost. The net information value subtracts the supplied study cost. Both quantities should be reported because they answer different commissioning questions.

The useful-signal fixture has a conditional policy that differs from the current Stage choice. Commit is selected after the positive signal and Hold after the negative signal. The gross improvement arises from making a more favourable response in each branch under the stated posterior. This is not simply a reward for a more accurate forecast. The forecast matters through the action consequences. A signal with substantial discriminatory power can have little decision value if both resulting posteriors remain within the same action region.

The no-policy-change fixture demonstrates that mechanism. Its prior is already favourable enough that Commit remains best after both possible signals. The instrument changes the success probability, but the action does not change. The weighted conditional payoffs recover the current policy's expected payoff exactly in the current rational evaluation. Its information value is 0.00 USDk, with exact zero retained in the exported fraction record. This result follows from the unchanged policy rather than display rounding or a cancellation tolerance. The study cost remains positive in the illustrative comparison, so the purchase has negative net model value. A decision owner should still review whether omitted states or an unrepresented action could change this conclusion.

The false-positive-burden fixture reaches the same policy conclusion from a different direction. Both branches select Hold because neither posterior justifies a paid exposure under the supplied payoff matrix. This is not the same uncertainty configuration as the highly favourable no-change case. One starts from an action that already exploits the upside; the other starts from a robust refusal of the exposure. Both show why forecast improvement and decision improvement are distinct. Research value depends on where the posteriors fall relative to the action boundaries.

The expensive-signal fixture preserves the useful instrument and conditional actions while increasing the supplied acquisition price. Its gross information value remains the same, but its net value becomes negative. This clean comparison separates a methodological finding from a procurement recommendation. The instrument can be useful, and its reported performance can be credible, while its price is too high for this particular decision. A sponsor should not defend the purchase solely by pointing to the signal's gross benefit or a ceiling for perfect information.

The restricted-action fixture offers a third route to a poor purchase. The instrument remains unchanged, but the institution cannot take Commit. The positive signal can still improve the response within the remaining menu, yet the improvement is too small to justify the supplied price. This case is a practical warning against attributing all value to the evidence provider. The client action architecture determines how much useful evidence can become an improved outcome. A successful study delivery does not guarantee an executable response with enough value to cover the commission.

The purchase rule is conditional on including the relevant study cost. The app accepts a single aggregate amount. If the report is delayed, internal analysis is extensive or an operating trial disrupts service, those consequences need their own evidence and valuation. They may be included in a documented aggregate monetary cost if appropriate, or imposed as separate constraints. The engine does not calculate them automatically. A positive net information value based only on a vendor invoice is not a complete commissioning appraisal when other material costs remain unresolved.

The gross value provides a maximum justified purchase price within this fixture's assumptions, but that statement needs care. It is the price at which net expected monetary value reaches the comparison boundary, not a recommended bid or an objective market valuation of the study. Different institutions can have different payoffs and response menus, so the same instrument can support different willingness to pay. A procurement team should retain commercial negotiation, deliverable quality and contractual scope as distinct issues. The decision value is an internal ceiling for a specified use, not evidence of the provider's cost.

The gross signal value should not exceed the perfect-information value when the state, menu and payoff perimeter are fixed. This relationship is a useful reproducibility check and a semantic check on the study design. A claimed signal benefit above the ceiling can indicate a changed menu, inconsistent baseline, overlapping benefit or an implementation mechanism that has been added without disclosure. The right response is to inspect the perimeter rather than force the result into an expected relationship. The paper's frozen fixtures retain each component for that inspection.

There is a subtle distinction between an information purchase and an action commitment. Approving the study means accepting an expected-value rationale for learning under the declared decision. It does not authorise whichever action an eventual result recommends. The committee may need a separate review of feasibility, rights, liquidity and

contractual conditions after the evidence arrives. Those conditions should be anticipated in the action charter so that the study does not depend on an implausible automatic commitment. A conditional policy can be a planning rule without being an unconditional execution mandate.

The study specification should preserve the expected action changes. If the final instrument addresses a different state, has different conditional performance or omits a branch, the original information-value calculation may no longer justify the commission. Delivery review should therefore compare the actual evidence process with the commissioned likelihood model and response requirements. The paper does not invent field performance for any instrument. It supplies a structure for checking whether a real deliverable answers the decision that was used to justify its cost.

An uncertain instrument can require a richer information-value model. The app treats sensitivity and specificity as fixed inputs, so it does not integrate uncertainty about those rates or evaluate an adaptive study design. A team may run explicitly varied fixtures to see how the purchase result changes, but that is a scenario comparison rather than a posterior distribution over study effectiveness. If performance uncertainty is consequential, the commissioning package should state the missing integration and use a suitable method. More plotted scenarios do not automatically become an empirical confidence interval.

The case slate supplies a purchase-failure classifier. No-policy-change fails through action invariance; expensive-signal fails through acquisition burden; restricted-action fails through the missing response; false-positive-burden exposes a signal that does not move the policy under its supplied account. A hypothetical procurement team should therefore identify the failure route before requesting an improved proposal. A discount can address price, but cannot make an absent response feasible. Better signal discrimination can affect a policy boundary, but cannot create authority. A redesigned action menu can change attainable value while leaving the instrument unchanged. Classifying the route prevents a generic request for more accurate evidence from missing the decision defect.

The purchase record should also distinguish evidence acceptance from action selection. A hypothetical instrument can deliver exactly the contracted result and still lead to Hold under the predeclared policy. That is a fulfilled commission, not a failed study. An instrument can deliver a persuasive result outside its accepted scope and leave the conditional policy unusable; that is an evidence problem even if the eventual action succeeds. An accepted result followed by an unavailable response creates an implementation problem. Recording these dispositions keeps the

gross-value calculation tied to the assumed policy and provides a concrete basis for later contract and learning reviews, without claiming that a single realised outcome measures the expected value of the study.

Evidence and scope

DOE links measurement plans to defined baselines, methods and reporting requirements for performance verification.

Evidence: Measurement and Verification Activities Required in the Energy Savings Performance Contract Process

Gross and net signal value

Conditional-policy expected payoff = sum over possible signals of branch probability × best conditional expected payoff. Gross signal value = conditional-policy expected payoff minus best current expected payoff. Net information value = gross signal value minus the separately supplied study cost. The comparisons retain the same actions and payoff perimeter.

Illustrative calculation. Assumptions and sensitivity are published with the model.

Illustrative purchase decisions across distinct mechanisms

All amounts are in USDk. The unchanged-policy information value is exactly zero under the current rational evaluation. Complete JSON retains the exact fractions and their numeric display values.

Illustrative purchase decisions across distinct mechanisms				
Fixture	Gross signal value	Study cost	Net value	Mechanism
Useful signal	33.20	5.00	28.20	Conditional actions improve the policy
No policy change	0.00	5.00	-5.00	Both branches still choose Commit
Expensive signal	33.20	80.00	-46.80	Same useful signal, greater price
False positive burden	0.00	4.00	-4.00	Both branches still choose Hold
Restricted action	3.45	5.00	-1.55	A valuable response is unavailable

Buy the conditional action improvement

Approve a decision-value commission only after comparing its expected conditional policy with the best current action and including the relevant cost. Keep gross usefulness separate from affordability. Reassess if instrument performance, action availability or the opportunity changes. Record any separate mandatory or exploratory purpose without claiming that the binary purchase calculation measures it.

1. Compare against the best current feasible action.
2. Inspect all branches of the proposed policy.
3. Place the complete relevant cost in the perimeter.
4. Preserve the ceiling and limitations beside the net result.

A negative immediate result can coexist with broader benefits

A study may support several future decisions, staff capability or a public evidence obligation. Those benefits can justify a broader commission when they are specified and reviewed. They should not be attached as vague rescue terms to a negative current-decision result. Define the additional use, avoid overlapping benefits and use a method appropriate to its horizon and action architecture.

The instrument-specific value assumes fixed performance

The engine does not infer performance from data or integrate uncertainty in its conditional likelihoods. Its net value uses a supplied aggregate cost and the fixed binary decision. Scenario comparisons can challenge these inputs but do not establish empirical instrument effectiveness, a sequential stopping rule or a research programme valuation.

Decision implications

- A signal creates value through the conditional policy it enables.
- Gross usefulness can coexist with negative net purchase value.
- An unchanged policy can make an informative signal economically redundant for the stated decision.

MECHANISMS AND WORKED STUDIES

09

The decision boundary matters more than average confidence

A sensitivity and falsifier map

Find the action switch that evidence can affect

A decision boundary is a point at which a change in the information available to the institution changes its preferred feasible action. The boundary depends on the payoff matrix and objective. It is not a universal

probability at which a project becomes acceptable. In the binary expected-value model, each action's expected payoff changes linearly with the prior. Comparing those lines shows where the optimiser can move from one action to another. This gives a concrete reason to ask for better information: the current evidence may leave the decision near a consequential switch.

The app reports a sampled prior policy and the intersections of each pair of action value lines. These are related but distinct outputs. A pairwise intersection tells the reader where those two actions have equal expected payoff. It does not guarantee that either is best among the full menu at that point. Another action can have greater value. The report therefore retains the app's best-of-all-actions reading at each intersection. Treating every pairwise crossing as an actual policy switch would overstate the information a simple threshold table supplies.

In the useful-signal matrix, the Commit and Hold lines cross at a supplied computed prior where Stage is better than both. This is an especially useful check. A diagram showing only Commit and Hold would suggest a commitment boundary that is not the boundary of the complete decision. The robust middle action changes the policy architecture. The committee should review the complete menu before using an attractive threshold to justify a full commitment, and the analyst should show which comparisons are inactive in the global optimum.

The sampled prior sweep is a visual guide, not an exact list of every transition. Its points are supplied by the canonical engine at fixed prior increments. A policy switch can occur between them. The pairwise intersections offer more precise candidate checks under the linear payoff structure, while the best-action field distinguishes an actual upper-envelope comparison from an inactive crossing. For a richer nonlinear or constrained model, a sampled sweep would need additional analysis. The paper does not claim that its app solves every possible decision boundary.

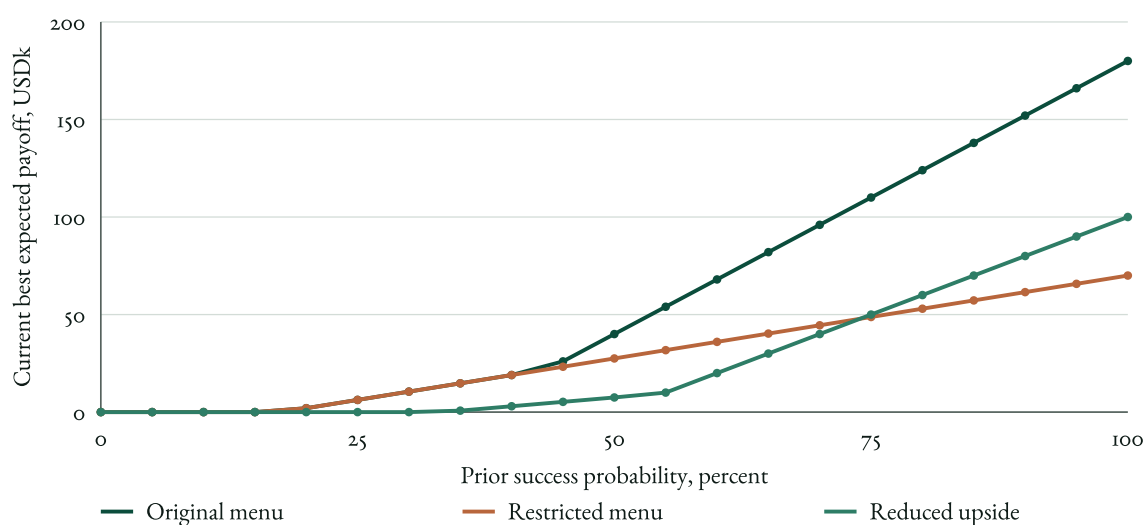
Information matters when it can move a posterior into a different action region. A signal that narrows uncertainty but leaves both branches in the same region has no algebraic action benefit in this fixed model. A less precise signal can be useful if it moves a meaningful branch across a boundary and supports a different response. This does not imply that instrument quality is irrelevant. It means quality should be judged against the decision architecture and the consequence of each branch, not against a generic target for forecast accuracy divorced from the available actions.

The useful and no-policy-change studies share the same payoff matrix and likelihoods but use different priors. In one, the conditional posteriors select different actions; in the other, both remain in the Commit region. The contrast demonstrates that an instrument does not possess a fixed commercial information value independent of the client's current evidence. The same instrument can support a valuable commission at one point and an economically redundant commission at another. The commissioning record should therefore be dated and should preserve the current prior rather than cite an earlier project's value.

Illustrative calculation. Assumptions and sensitivity are published with the model.

The response menu changes the current policy account

The plotted values are the canonical sampled best expected payoffs at alternative priors. They compare the original, restricted and reduced-upside menus before buying a signal. Connecting segments aid reading between sampled points; they do not publish an EVSI sweep or a probability distribution over priors.



Original deterministic illustration from frozen current-policy output. Every point binds to its named fixture and actual output pointer. The acquisition cost and signal likelihoods do not enter this current-policy sweep.

The false-positive-burden case also shows why the favourable branch should be evaluated against the actual boundary. A positive signal can materially increase the success probability and still leave Hold preferable. Describing it as a strong improvement in belief can be correct while describing it as a reason to commit can be wrong. The committee should ask what action becomes feasible and preferable at the posterior, not only whether the posterior is larger than the prior. This avoids confusing a direction of update with a sufficient decision condition.

Boundary proximity does not automatically mean more research should be purchased. The signal's cost, attainable performance, timing and available response all still matter. A decision near a crossing may involve small monetary differences or an unresolved nonmonetary constraint. A costly study can be disproportionate even if it changes the selected action. Conversely, a meaningful downside can justify improving evidence when a small probability change has large consequences. The signal-value calculation integrates the supplied payoffs and branch masses; the boundary analysis explains where its action improvements originate.

The committee should distinguish uncertain parameters from sensitivity ranges. A sweep across priors shows what the decision would do if different values were supplied. It does not establish that every prior in the sweep is equally plausible or that the current prior has a uniform distribution. A separate evidence account should describe the plausible range and its basis. If the input owner cannot support that range, the sweep remains a mechanism test. Its value is to identify where additional evidence could matter, not to manufacture an empirical probability distribution from a convenient horizontal axis.

The same applies to a payoff sensitivity. Changing a failure exposure can move the boundary without changing the instrument. A legal allocation, an operating safeguard or a smaller commitment may therefore reduce the need for research by making the action more robust. This is an institutional choice rather than an evidence forecast. A team should compare feasible ways to improve the decision: buy better information, alter the action, reduce a consequence or accept an explicitly bounded exposure. The model can compare supplied revised matrices, but the feasibility and cost of those changes require separate evidence.

An intersection should not be treated as a sharp empirical fact when the input cells are rough. The algebra can produce an exact point from approximate values. Reporting that point with many decimals may obscure the uncertainty in the boundary itself. The paper displays the canonical outputs at a declared precision for reproducibility, and the tables identify their illustrative status. An actual decision dossier should preserve the computational point while explaining the evidence range and practical conditions around it. Precision of calculation is different from precision of the institution's knowledge.

The response rule should also handle a result close to the boundary. The app provides a deterministic tie-break for reproducible output. A committee may instead request an admissibility check, choose a more reversible action or require a defined margin before a consequential commitment. Those conditions should be agreed before receiving the

signal where feasible. Otherwise a sponsor can selectively invoke the boundary when it favours commitment and dismiss it when it favours holding. A predeclared rule makes that inconsistency reviewable.

Policy charts should show what is plotted. The engine's prior sweep displays the best expected payoff, which is the upper envelope of the supplied actions at each sampled prior. It does not display signal value at each prior or the chance that a threshold will be crossed. Those would require additional computations. A report that titles the existing sweep information value versus prior would misstate the output. The graphical exhibit should retain its axes, units, action scope and selected canonical pointers so that a reader can distinguish a policy picture from an acquisition-value picture.

A boundary enquiry should test the action envelope, not merely a pairwise intersection. If the crossing actions are both below Stage, moving a posterior across their intersection does not change the selected policy. The useful enquiry is whether the evidence can reach a region where a different admissible action becomes best. An uncertain payoff can move that region as well as an uncertain prior, so retain the reason for each sensitivity. A hypothetical team targeting the inactive Commit and Hold crossing would be buying precision at a point that does not govern the current menu. The existing envelope and branch account expose that error without assigning a probability to parameter uncertainty or pretending that the sampled sweep computes acquisition value.

Evidence and scope

The Magenta Book connects evaluation planning with decision needs and the use of findings.

Evidence: Magenta Book: Central Government guidance on evaluation

Illustrative calculation. Assumptions and sensitivity are published with the model.

Illustrative pairwise intersections and the complete-menu winner

The table preserves the canonical distinction between pairwise equality and the globally preferred action.

Illustrative pairwise intersections and the complete-menu winner			
Action A	Action B	Prior at equality, percent	Best among all actions
Commit	Stage	43.59	Commit
Commit	Hold	35.71	Stage
Stage	Hold	17.65	Hold

Research around an active boundary

Check the full action menu before interpreting a crossing. Compare possible posteriors with the complete-menu action regions, retain the evidence basis for a plausible prior range, and explain whether the proposed study can move a branch into a materially different response. Separate a sampled policy plot from a signal-value calculation. Agree treatment of near-boundary findings before the result arrives.

1. Identify active and inactive pairwise comparisons.
2. Map the possible posteriors to feasible action regions.
3. Challenge uncertainty in the payoff boundary.
4. Record any required decision margin.

A robust choice may still deserve monitoring

A current action that remains best across the supported range can reduce the case for another immediate study. It does not remove the need to watch for a changed opportunity, new constraint or evidence outside that range. Define a monitoring trigger tied to a material change in the decision perimeter. This preserves robustness without treating the current policy as permanently immune to new information.

The policy sweep is sampled and does not assign plausibility

The engine reports a finite prior sweep and pairwise linear intersections. It does not compute an empirical distribution of priors, uncertainty intervals for thresholds or a signal-value curve across the sweep. Its best-action reading must be interpreted with the complete supplied menu. Broader boundary claims require a model that actually calculates them.

Decision implications

- A pairwise intersection may be inactive in the complete action menu.
- Useful information moves a consequential posterior across an actionable boundary.
- A sampled prior policy is different from a signal-value curve.

MECHANISMS AND WORKED STUDIES

10 Delay is an information cost

A research timing decision framework

Evidence needs time to become a feasible response

Evidence has to arrive while the decision can still use it. A study that produces a credible result after the commitment becomes irreversible can serve learning or accountability, but it cannot improve the earlier action through the conditional policy used to justify its purchase. The commissioning charter should therefore identify the latest usable evidence

date, the review interval and the action deadline separately. A delivery date alone is insufficient when the institution also needs time to validate the result, convene an authority and execute its response.

Delay can change both information value and the underlying opportunity. The current action menu may narrow as rights expire or a delivery slot disappears. Payoffs may change as prices, capacity or costs move. The state likelihoods may change because new evidence arrives elsewhere. A dated appraisal should distinguish these mechanisms rather than collect them under a generic discount factor. Each mechanism has a different consequence for the study: it may reduce the opportunity, alter the current benchmark, change the feasible branch response or invalidate the instrument's use context.

The Decision lab does not contain a time-delay input or a cash-flow discounting engine. The paper's reduced-upside case uses a separately supplied payoff matrix to illustrate an opportunity that has become less attractive. It deliberately does not claim that a particular study duration caused the reduced payoffs or calculate a rate of value loss over time. This distinction is necessary because an illustrative matrix comparison can reveal sensitivity to changed opportunity value while leaving the actual schedule-to-payoff relationship entirely unresolved.

The reduced-upside fixture retains the useful signal's prior and likelihoods. Its state payoffs are different, so both the current value and the conditional-policy value change. The signal can still add value within that revised decision, but its monetary gain differs from the original commission. A team that continues to quote the original gross benefit after the opportunity changes is using an outdated perimeter. The frozen pair of fixtures lets a reader see the comparison and the explicit limitation without implying an empirical estimate of delay cost.

An actual schedule review should start from the event that makes a response unavailable. It may be a contract signature, a capital commitment, a regulatory submission, a physical installation or a service obligation. The institution should identify which parts of the decision are reversible before that event and which are already fixed. Research can still influence a remaining design choice even after another part of the project is committed. The charter should specify the residual action rather than suggest that the report can reverse the entire programme.

There is also a distinction between evidence delivery and evidence usability. A technically complete file may need methodological checking, interpretation by operating staff, a governance review or a legal assessment. Those steps take resources and can reveal that the result is

insufficient for the intended response. A study contract that measures only file delivery can therefore leave the decision unsupported. The completed commissioning timeline should include acceptance of the evidence and authorisation of the action, with owners responsible for each transition. This is a governance model rather than an estimated schedule in the binary app.

The institution should ask what happens while it waits. Holding a decision may preserve capital but impose an operating cost, lose a customer or expose a facility to an unresolved risk. Acting provisionally may create a sunk commitment that changes later payoffs. A temporary safeguard may be available at a price. These interim actions belong in an appropriate sequential account when they determine the choice. They should not be omitted because the research price appears small. The app's single aggregate cost cannot reconstruct an unrecorded waiting strategy.

The schedule can also create selection bias in how evidence is used. A favourable preliminary result may be circulated early to preserve a commitment date while an adverse result waits for more checking. The apparent decision then becomes conditional on the organisation's reporting incentives rather than only the instrument. A predeclared release and review rule should treat material findings consistently. The paper does not assert that any actual institution follows this pattern. It identifies a plausible commissioning failure that a decision record can prevent by naming the review path in advance.

Delay risk should be assigned to someone who can manage it. A researcher may control data collection but not access to a site. A sponsor may control committee scheduling but not a regulator's response. A procurement team may control contracting but not an operating trial window. A single promised delivery date can conceal these interfaces. The decision package should separate dependencies, indicate the evidence needed to accept each one and define what happens if the latest useful date is missed. A report that arrives too late should not be retroactively credited with changing an action already taken.

Research progress can warrant an interim review without producing a valid partial signal. A team may discover that the original test will not meet its planned conditions or that the intended population cannot be accessed. This is evidence about the commission's feasibility, not necessarily evidence about the opportunity's success state. The institution should reassess the remaining cost and decision purpose. The binary engine does not calculate adaptive continuation value. It can support a revised fixture after assumptions change, provided the revision is explicitly documented and not portrayed as a preplanned sequential calculation.

The purchase comparison should use costs that remain avoidable at the review date. A fee already irreversibly paid does not become recoverable because the study now has negative remaining value. Conversely, continuing work can have incremental costs and benefits distinct from the initial commission. A practical stopping review should record the remaining deliverable, cost, action window and alternatives. This avoids using either the original sunk price or the desire to defend it as the sole reason to continue. The later stopping chapter formalises this as a record of the current decision, not a moral judgement on the initial team.

An opportunity may improve during the study. Delay is not automatically harmful. A price change, a new operating capability or an additional feasible action can make the signal more valuable. The institution should therefore monitor the decision perimeter without assuming a fixed direction of change. A revision should be driven by evidence and a documented event, not by adjusting payoffs to preserve the original purchase recommendation. The raw fixture history gives reviewers a way to distinguish a changed opportunity from an opportunistic change in the appraisal's assumptions.

A schedule dispute should be classified by what becomes unavailable. In a hypothetical commission, the instrument reports on time but the receiving team cannot review it before a resource reservation expires. The defect lies between delivered evidence and usable response, not in laboratory turnaround. If the report itself arrives after the response deadline, the supplier's delivery route is relevant. If the deadline remains open but the opportunity's payoff changes, the consequence account is relevant. Protected or statutory conditions require their own authority review; an arbitrary delay penalty cannot restore an inadmissible response. This classification locates the changed object before a static monetary comparison is rerun.

Back-plan from the last reversible response event through interpretation, review and evidence delivery. A hypothetical early draft is useful only if its status permits the intended decision; an unverified preview cannot be assigned the likelihoods of an accepted final report. At each interface, state the fallback if the required object is absent: retain the currently feasible action, release an expiring reservation where permitted, or escalate the decision to a different commission. The resulting schedule is a sequence of usable evidence objects rather than a list of publication dates. It reveals where a timing intervention can preserve value while leaving the binary app's absence of a dated delay calculation explicit.

Evidence and scope

GAO scheduling guidance requires logical dependencies and a credible basis for completion dates.

Evidence: [Schedule Assessment Guide: Best Practices for Project Schedules](#)

Illustrative calculation. Assumptions and sensitivity are published with the model.

Illustrative changed opportunity: supplied upside comparison

The reduced matrix is assumed. The engine does not calculate a time interval, discount rate or causal delay effect.

Illustrative changed opportunity: supplied upside comparison			
Action	Original success, USDk	Revised success, USDk	Failure convention
Commit	180.00	100.00	Failure payoff unchanged
Stage	70.00	30.00	Failure payoff unchanged
Hold	0.00	0.00	Baseline unchanged

Illustrative calculation. Assumptions and sensitivity are published with the model.

Illustrative revised decision values

The change is a payoff sensitivity, not an estimate of research delay cost.

Illustrative revised decision values			
Fixture	Current value, USDk	Gross signal value, USDk	Net value, USDk
Original opportunity	19.00	33.20	28.20
Reduced upside	3.00	22.00	17.00

Completed illustrative evidence-to-action timeline

Relative events are used because no actual project schedule is asserted.

Completed illustrative evidence-to-action timeline			
Event	Owner	Acceptance condition	Fallback if unavailable
Commission approved	Committee	Decision and cost perimeter accepted	Revise or decline commission
Evidence access confirmed	Study lead	Relevant population and conditions available	Review instrument feasibility
Result delivered	Research provider	Complete commissioned evidence object	Assess remaining usable decision
Method checked	Independent reviewer	Material limitations understood	Withhold unsupported branch response
Action authorised	Delegated authority	Feasible response within remaining window	Choose admissible fallback
Decision window closes	Action owner	Irreversible boundary recorded	Use result for declared later purpose only

Approve a usable evidence window

Identify the event that makes each response unavailable, include review and execution time, and name dependency owners. Reassess the current decision when payoffs, rights or evidence change. A reduced-payoff fixture can expose sensitivity but should not be presented as a calculated delay effect. Record the fallback if evidence cannot be accepted before the usable window closes.

1. Separate delivery, acceptance and action dates.
2. Identify irreversible boundaries.
3. Review changes to the opportunity perimeter.
4. Define a fallback before the study is commissioned.

An expiring decision should not force a rushed study

A deadline can create pressure to accept weak evidence. The institution can choose a smaller feasible action, an explicit hold or another safeguard rather than treating an incomplete result as decisive. The cost of that alternative should be examined, but the deadline does not validate the instrument. A commission that cannot meet the needed standard in time should have its purpose or action architecture revised.

Timing mechanisms are conceptual here

The engine has no dated research schedule, discount calculation, adaptive continuation model or action-window logic. The reduced-upside fixture is a separately supplied matrix. Real claims about delay cost require a supported schedule-to-consequence relationship and appropriate calculation. The qualitative timeline identifies that evidence obligation rather than implying it has already been met.

Decision implications

- A usable evidence date includes acceptance and action authorisation.
- Changed opportunity payoffs require a new comparison.
- A supplied reduced-upside matrix does not calculate the causal cost of delay.

MECHANISMS AND WORKED STUDIES

11 Information can fail through an implementation gap

An actionability acceptance worksheet

Close the gap between a posterior and an executable action

The information-value calculation assumes that the institution can select and execute the action recommended in each branch. This assumption is often more demanding than the arithmetic. A favourable posterior may require a response that exceeds delegated authority, depends on another party or consumes a resource already committed elsewhere. A negative posterior may require withdrawal from an obligation that cannot be abandoned. The institution should review these conditions before purchasing a study whose value relies on them. Evidence becomes useful through a response path, not through an attractive branch label alone.

The restricted-action fixture provides a clean numerical test of this implementation gap. It removes Commit while retaining the prior, likelihoods and the remaining action payoffs. The instrument still changes beliefs and can still alter the action between Stage and Hold. However, its gross improvement falls because the high-upside response is unavailable. The same price now exceeds the decision benefit. This does not estimate the value of obtaining approval for Commit. It shows that an information commission justified with an unexecutable menu can be economically misleading.

The review should distinguish authority from capability. A committee may have the power to approve a redesign but lack the technical team to implement it. An operator may have the capability to change a process but lack permission under a service agreement. A sponsor may possess both and still depend on an external interface. These are different prerequisites with different owners and evidence. A generic statement that management

will act does not establish any of them. The action register should identify the concrete transition from a reviewed signal to an operating or commercial change.

Implementation can also require another information decision. A positive technical result may justify a detailed design study before a commitment, rather than immediate construction. A negative result may justify a diagnosis of the failure mechanism rather than abandonment. The simple app cannot infer those sequential responses from an action name. If the committee's actual next choice is another commission, that action should have its own perimeter and analysis. Describing a broad future strategy as a single action is acceptable only when the supplied aggregate consequences and limitations make the compression explicit.

An institution should avoid building a decision rule around a response it cannot credibly commit to consider. If the sponsor has already decided that an adverse signal will be ignored, the purported benefit of avoiding failure exposure is absent from the actual policy. A model can still show what a rational response would do under its objective, but it should not claim that this is the organisation's implemented conditional policy. A credible charter needs a predeclared review process and an authority capable of changing the commitment when the evidence warrants it.

The evidence provider also needs a clear boundary. A research team can deliver a methodologically sound result without controlling the client's response. The study contract should specify the evidence object, method, scope and limitations, rather than promise the success of a later operating decision. The client should own the action architecture and acceptance process. This division prevents a provider from being credited with unimplemented benefits or blamed for a response it was never authorised to control. It also makes the commissioning rationale more honest about where value must be created.

The handoff should preserve the meaning of the signal. Operating staff need to know which state the posterior describes, the conditions under which the instrument was evaluated, and the actions that were reviewed. A summary stating that the test passed can conceal conditional uncertainty and unrepresented outcomes. The branch record should travel with the decision package, including the adverse response and the circumstances that require escalation. The app's export supports a reproducible numerical object; the institution must still organise the human handoff that makes it usable.

An action may remain feasible but lose its original payoff when implementation is delayed or incomplete. The fixture assumes the supplied response consequences occur under the selected action. In practice, a partial redesign, missed interface or weak operating adoption can change those consequences. The simple engine does not apply an implementation success probability or calculate adoption cost. A team can supply revised aggregate payoffs only when it explains their basis. If implementation uncertainty is material and distinct from the opportunity state, a broader state or sequential model may be needed.

The response path should contain acceptance criteria. For a full commitment, this may mean confirming a defined technical condition and remaining obligations. For staging, it may mean identifying what the initial phase authorises and what remains contingent. For holding, it may mean checking service duties and opportunity expiry. These criteria should be relevant to the action, proportionate to the consequence and owned by someone who can verify them. They are not universal checklist items; the institution should define them from its actual mandate and evidence rather than copy a hypothetical worksheet without adaptation.

Governance can alter information value without altering signal accuracy. A clearer delegated scope, a feasible fallback or an agreed response review can make an existing study more usable. This does not mean governance automatically creates monetary value. The action consequences and costs need their own support. It means that the evidence commission and implementation architecture should be designed together. Buying a more accurate signal while leaving response authority unresolved can spend resources on precision that the institution still cannot exploit.

The implementation record should also identify who may reject a response despite its numerical attractiveness. A protected interest, unresolved obligation or material safety condition can make an action inadmissible. The model's expected-value optimum is not a defence against such a boundary. The appropriate authority should review the constraint and the supporting evidence. This preserves the distinction between a financial comparison and an institutional permission. The paper's hypothetical roles illustrate that division without inventing a legal conclusion or an actual client's approval process.

After the signal arrives, the committee should compare the proposed response with the predeclared rule. A justified change can occur because new evidence, a changed opportunity or an uncovered constraint alters the decision. The record should describe that reason. An unexplained deviation can indicate that the original study value was overstated or that the action architecture was never accepted. Reviewing deviations is useful

for improving future commissions, provided the institution does not infer empirical performance rates from a small selected set of cases without an appropriate universe and method.

Test the response path with a hypothetical favourable result delivered while Commit is not executable. If permission is absent, obtain the appropriate authority decision; if a compatible resource is absent, resolve capacity; if the result is outside the accepted population, return to instrument scope. These failures can look identical in a dashboard that simply reports no action taken, yet need different repairs. Stage or Hold may remain usable fallbacks under the supplied menu. Record the selected fallback and the reason the intended branch could not advance. A later favourable outcome does not retroactively make Commit available at the decision date, and a later adverse outcome does not identify which interface failed.

An implementation exercise should stop at the first unmet prerequisite. In a hypothetical paper exercise, branch recipients receive the accepted signal, identify the response, obtain the required authority and reserve its operating resources. If reservation fails, continuing to sign a nominal response instruction demonstrates paperwork rather than execution. Retain the failed interface and test a permitted fallback instead. Passing the exercise establishes the interfaces actually exercised, while field performance and payoffs remain unvalidated. This makes the response worksheet a discriminating instrument and shows why the restricted-action comparison is an implementation sensitivity, rather than a calculation of the value of persuading an authority to approve an action.

Evidence and scope

NIST risk framing identifies assumptions, constraints, tolerance and institutional priorities before selecting responses.

Evidence: Managing Information Security Risk: Organization, Mission, and Information System View

Completed illustrative response-path register

The roles and acceptance conditions are hypothetical and require local adaptation.

Completed illustrative response-path register			
Action	Authority owner	Capability condition	Acceptance before response
Commit	Delegated investment committee	Full-scope operating delivery path	All commitment constraints resolved
Stage	Programme sponsor within phase scope	Bounded first-phase execution team	Aggregate strategy and later approvals understood
Hold	Accountable action owner	Ability to preserve an admissible baseline	Opportunity and service duties checked
Escalate unrepresented result	Independent method and mandate reviewers	Access to evidence and action owners	Model boundary acknowledged before new response

Illustrative calculation. Assumptions and sensitivity are published with the model.

Illustrative implementation sensitivity

The removed action reduces the value that the same instrument can create.

Illustrative implementation sensitivity				
Menu	Positive response	Negative response	Gross value, USDk	Net value, USDk
Unrestricted	Commit	Hold	33.20	28.20
Commit unavailable	Stage	Hold	3.45	-1.55

Certify a response path with the commission

Confirm authority, capability, interfaces and acceptance conditions for each branch response. Exclude unavailable actions from the operative menu. Preserve the signal interpretation in the handoff and record justified deviations from the predeclared response. Separate evidence delivery, implemented action and later outcome when evaluating the commission.

1. Name the branch action owner.
2. Check authority and capability separately.
3. Resolve dependencies before relying on their payoffs.
4. Retain an escalation rule for model boundaries.

Conditional approval can be sufficient at an early stage

A committee need not resolve every implementation detail before any research. It can approve a bounded commission whose next response is an additional feasibility decision. The operative action menu should then describe that next decision and its supplied consequences, rather than pretend the full commitment is already executable. Research scope should match the actual authority available at the current stage.

The engine assumes the selected response occurs

The app has no adoption, authority, dependency or implementation-failure calculation. The restricted-action fixture changes the menu explicitly, while the qualitative worksheets expose other premises that may need a broader model. A numerical branch choice does not establish permission to execute or predict field performance.

Decision implications

- Signal accuracy does not establish response authority or capability.
- An unexecutable action can materially overstate information value.
- Evidence delivery, implemented action and subsequent outcome should be recorded separately.

MECHANISMS AND WORKED STUDIES**12****Risk aversion changes the decision criterion***A method-selection and limitation framework***The monetary objective has institutional boundaries**

The expected monetary rule used in the worked studies is a declared objective. It is not a universal account of institutional preference. A committee may care about the distribution of outcomes, liquidity in an adverse state, a mandate constraint or an exposure that cannot be diversified. Those concerns can make an action with a lower expected monetary payoff preferable under the institution's actual decision rule. The analyst should identify the disagreement as an objective or admissibility question instead of insisting that correct expected-value arithmetic settles it.

Risk aversion concerns how consequences are valued, not merely how uncertain they are. A probability expresses a belief about a represented state. A utility function would express a preference over its consequences.

Changing the prior to make a risky action less attractive is an inappropriate way to insert risk aversion when the belief itself has not changed. The institution should preserve the probability evidence and state the preference or constraint separately. This keeps the model's epistemic assumptions from becoming an instrument for achieving a desired recommendation.

The binary app has no utility-function input. It therefore cannot calculate expected utility, a certainty equivalent or an optimal action under a chosen nonlinear preference. The paper discusses these methods conceptually and uses the monetary outputs as a reference comparison only. A real utility analysis would need a justified consequence measure, a preference elicitation process and appropriate calculation. An analyst should not apply an arbitrary haircut to the monetary result and call that a calibrated risk-aversion adjustment. Such a change is another supplied payoff assumption unless its method is actually defined and supported.

Constraints provide a different way to represent some institutional concerns. An organisation may rule out an action that exposes it to an unacceptable service interruption or breaches a mandate, regardless of average payoff. This is an admissibility condition rather than a smooth preference penalty. The action should be excluded from the operative menu or reviewed with a model that explicitly handles the constraint. The restricted-action fixture illustrates the effect of removing a response but does not establish why any real response should be prohibited. That decision belongs to the relevant authority and evidence.

The distinction matters for information value. An instrument that enables a high monetary payoff through an inadmissible action can have an overstated apparent benefit. An instrument that reduces exposure under a utility objective can have a different value from its risk-neutral monetary gain. The current paper cannot infer either value from the existing engine. It can identify which assumption would change and what additional method is required. The committee should avoid translating an attractive monetary EVSI directly into a research budget when its actual objective has not been reconciled with the model.

Expected regret offers another transparent reading of the same supplied matrix. The engine calculates state-specific regrets and their probability-weighted expectation for each action. The current monetary optimum minimises expected regret in this fixed menu because the perfectly informed benchmark is the same across actions. This equivalence does not establish a maximum-regret rule or a psychological measure of regret. A maximum-regret decision would focus on a different aggregation of state

consequences and need an explicitly declared method. The state-regret table allows readers to inspect the inputs without pretending that the app has selected that alternative rule.

The illustrative regret table reveals different exposure patterns. Commit gives up no payoff relative to the best action in success, but has a large failure-state regret. Hold avoids failure-state regret while giving up the success upside. Stage divides the sacrifice differently. These are state-contingent consequences of the supplied matrix. A committee can use them to discuss its mandate and tolerable exposure, but it should not infer a universal preference for the middle action or the largest success payoff. The selected rule and evidence perimeter determine what the table means for the institution.

Risk capacity also differs from risk preference. An institution might be willing to accept a downside but unable to fund the cash consequence when it occurs. A present-value payoff can conceal that timing problem. Another institution may have capacity but a mandate that restricts the exposure. The binary payoff engine does not calculate dated liquidity, funding covenants or legal authority. A decision package should therefore retain a separate capacity and permission review when these conditions are relevant. The monetary result should not be used to imply that a funding path exists.

The committee should ask whether exposures are repeated and diversifiable. A linear monetary expectation can be a useful planning reference for a set of limited choices, but concentration and common causes can change the institution's capacity. The app does not model a portfolio or dependence across opportunities. If the research commission supports a single consequential commitment, an aggregate expectation should be accompanied by a review of the adverse state and response. The paper offers no empirical conclusion about how a particular institution should trade average return against concentration.

Preference elicitation requires its own governance. The person supplying a utility or constraint should have authority to represent the institution, and affected interests should be considered through the appropriate process. An analyst's convenient parameter is not automatically an institutional preference. The decision record should distinguish a provisional sensitivity assumption from an approved objective. If participants disagree, the report can show the consequences of alternative declared rules while preserving the unresolved judgement. Hiding the disagreement inside an adjusted payoff cell makes the result harder to challenge.

Research can itself help clarify preferences, but that is a different information problem. A workshop may reveal how decision owners value consequences or what constraints actually bind. This can improve the decision architecture without observing the opportunity's success state. The binary instrument fixtures do not calculate the value of eliciting institutional preferences. A team should describe that purpose directly and avoid borrowing the commercial state's EVSI as a budget justification for every governance or deliberation activity. The information object, decision and response mechanism should remain explicit.

A claim of prudent conservatism also needs a basis. Lowering all upside inputs, raising every downside and imposing a decision margin may duplicate the same concern several times. The result can become systematically cautious without a traceable risk rule. An evidence package should record which uncertainty is represented through probabilities, which consequence uncertainty is represented through payoff sensitivity, and which institutional preference is represented through an objective or constraint. This prevents a plausible cautionary label from replacing a coherent decision method.

Use an objective-change audit when the selected action differs from the monetary leader. A hypothetical committee can exclude an action because it exceeds a protected exposure, choose a supported utility rule because concentrated loss matters, or apply another authorised criterion. Each route changes a different part of the decision architecture. Removing the action changes the menu; utility changes valuation; another criterion adds a separate comparison. None should be hidden by rewriting a payoff cell until the monetary output matches the committee's preference. Preserve the monetary reference and identify the actual route. A remaining difference then explains the cost or consequence of the chosen rule instead of appearing to be an arithmetic inconsistency.

Preference disagreement also needs a different evidence request from payoff disagreement. If hypothetical reviewers agree on the consequences but disagree about accepting the downside, more precise cash estimates may not resolve their risk preference. If they disagree about whether the institution can fund that downside, the question is capacity and admissibility. If they disagree about the downside itself, the consequence evidence is relevant. The objective register should route each dispute accordingly. This is a practical use of the boundary around risk-neutral money: it prevents the team from purchasing a predictive signal to resolve a normative or financing question that the signal cannot answer.

Evidence and scope

The Orange Book integrates risk with objectives, governance and context-specific treatment decisions.

Evidence: The Orange Book: Management of Risk, Principles and Concepts

Illustrative calculation. Assumptions and sensitivity are published with the model.

Illustrative state and expected regrets

These are canonical regret outputs. The app does not calculate a maximum-regret policy or utility optimum.

Illustrative state and expected regrets			
Action	Success regret, USDk	Failure regret, USDk	Expected regret, USDk
Commit	0.00	100.00	60.00
Stage	110.00	15.00	53.00
Hold	180.00	0.00	72.00

Completed illustrative objective and capacity register

The additional reviews are identified explicitly and are not implied by the monetary engine.

Completed illustrative objective and capacity register		
Decision element	Declared treatment	Authority or evidence needed
Monetary comparison	Risk-neutral expected aggregate payoff	Committee accepts this reference objective
Admissibility	Unavailable actions removed	Mandate and response-owner review
Utility preferences	Unresolved; no utility computation	Approved preference method if consequential
Liquidity capacity	Outside aggregate payoff engine	Dated cash and funding analysis
Protected conditions	Reviewed as constraints	Appropriate accountable authority
Concentration	Outside the single-opportunity fixture	Portfolio and common-cause review

Keep belief, preference, capacity and permission distinct

Do not change a probability to hide a preference or treat a positive average as evidence of capacity. Approve an objective explicitly, remove inadmissible actions and name additional methods when utility, liquidity or concentration determine the choice. Retain the monetary reference and explain any final decision that rests on a broader institutional rule.

1. Approve the objective and its owner.
2. Preserve probability evidence independently.
3. Check capacity and permission separately.
4. State additional computations that remain required.

Several objectives can make the decision unwieldy

A committee can use a primary objective with explicit constraints to keep a decision tractable. It does not need to quantify every concern into a single synthetic score. The important condition is that consequential conflicts are visible and governed. Simplicity is useful when it preserves actual institutional decision boundaries; it becomes misleading when it removes them solely to obtain a clean optimiser result.

Alternative decision rules are conceptual

The app computes risk-neutral expected monetary values and the associated regrets for fixed binary states. It does not implement expected utility, certainty equivalents, maximum-regret optimisation, liquidity analysis or portfolio diversification. The worksheet identifies these boundaries and required evidence without claiming that they are reproduced model outputs.

Decision implications

- Probability, preference, capacity and permission should remain distinct.
- Expected regret is different from a maximum-regret decision rule.
- A committee can reject a monetary optimum while preserving a transparent explanation.

GOVERNANCE AND APPLICATION

13

Multiple uncertainties create conditional dependence

A qualitative dependency diagram and incremental-evidence challenge

Several reports may carry the same underlying evidence

Several pieces of evidence do not automatically provide several independent improvements in a decision. Signals may share data, a physical cause, an expert panel or a selection process. Their apparent agreement can therefore reflect the same information repeated through different channels. The binary engine evaluates a single supplied signal against a current prior. It has no joint-signal likelihood model. A report using it should not multiply several instrument claims or add separate information values as though dependence had been resolved.

The prior may already incorporate evidence similar to the proposed signal. If the committee has screened opportunities using an early technical result, a later test evaluated on the same screened population can overlap with what is already known. The correct current prior should condition on the existing information set, and the new likelihood should describe the additional signal in that context. The simple app accepts those numbers but does not derive the bridge. A provenance register should state what evidence has already entered the current belief and why the new instrument adds distinct information.

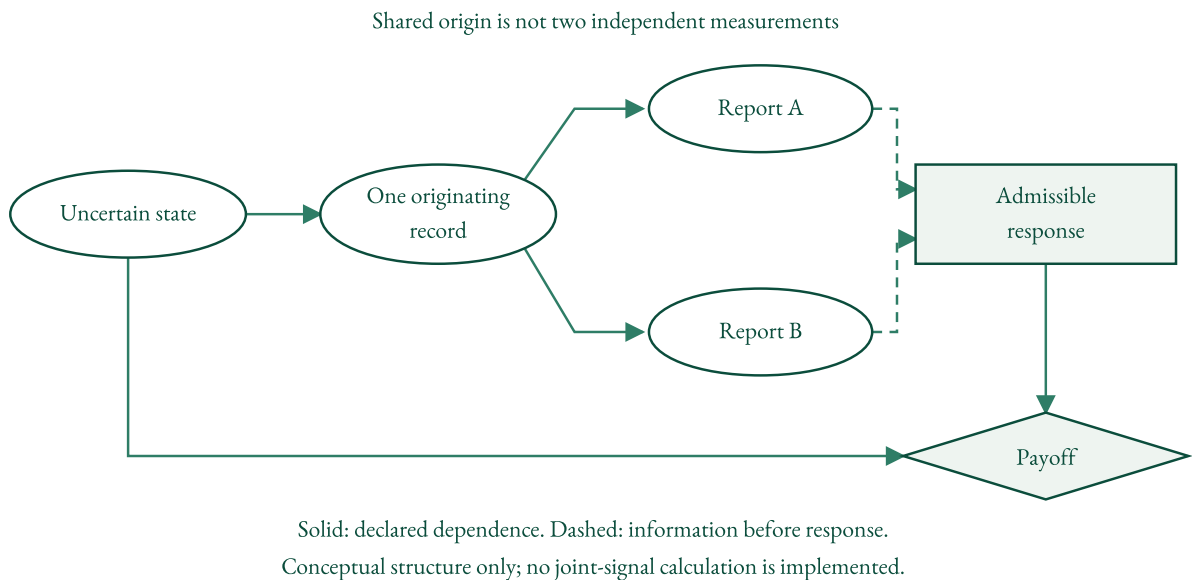
Shared data can create an especially subtle duplication. A market report and an internal forecast may both depend on the same survey, while a technical assessment and an engineering estimate may both use the same assumptions about operating conditions. Treating their agreement as independent confirmation can overstate confidence. The evidence map should trace inputs rather than count report titles. A committee should ask whether each source provides a new observation, a new model using old observations or a new interpretation of the same underlying evidence.

Shared expert judgement presents a similar issue. Several reports can cite the same specialist or institutional consensus. Their publication by different organisations does not necessarily make them independent evidence of the state. An elicitation process should record who participated, what information participants shared and how judgements

were aggregated. The current paper does not estimate correlations between experts or prescribe a universal aggregation rule. It recommends retaining the dependency question so that a numerical prior does not gain authority merely from repeated citation.

Shared origin is not independent confirmation

Original qualitative influence structure for a duplicated-origin evidence case. The reports share a parent record, so their likelihoods cannot be multiplied as independent new measurements without a justified joint model. Dashed edges identify information available before the response; the payoff depends on the state and chosen action. This diagram does not extend the binary-signal app into a joint or causal inference engine.



Signals can also depend on the state in a way that requires conditional modelling. Two instruments may both respond to an underlying defect or market condition, so their joint likelihood cannot be reconstructed from their individual marginal performance alone. A joint model needs evidence or assumptions about their combined outcomes given each state. The binary app supplies neither a joint table nor a graphical dependency network. A reader who wants to evaluate a sequence or combination of tests should use a method that actually represents those outcomes and preserve the assumptions for reproduction.

The independence assumption is not inherently unacceptable. It can be a useful simplification when supported or when a sensitivity study shows that dependence does not alter the choice. The problem is treating independence as the default merely because it makes arithmetic easier. A decision package should state why the assumption is reasonable for the intended use and how the recommendation changes if it fails. If those

checks have not been performed, the report should identify the missing analysis rather than present an independent-combination result as established evidence.

Adding separate information values can double count response benefits. If either of two signals alone would cause the committee to choose the same improved action, the second may have little incremental value after the first is known. The value of a signal should be evaluated against the best current action using the current information set. Once another signal has updated that set, the benchmark and branch probabilities can change. A sum of standalone EVSIs calculated from the same original prior does not generally represent the value of a research package.

The order of evidence acquisition can matter. An inexpensive early signal may determine whether an expensive study is still worthwhile. A costly signal may become redundant after an external result arrives. A later test may target a subgroup defined by an earlier result, changing its use population. These are sequential decisions with conditional acquisition costs and response options. The simple fixed-signal engine cannot calculate an adaptive programme from an action label. It can support explicitly revised snapshots, but a series of snapshots is not automatically a complete dynamic optimisation.

The charter should also distinguish repeated measurement from independent state evidence. Repeating an instrument can improve knowledge of measurement variability under appropriate conditions, but it may not resolve a shared systematic bias or an omitted state. A report that reduces a random error metric can still leave the decision uncertainty unchanged. The metrology account and the commercial state account need a supported bridge. This is why a precise measurement, a well-reproduced computation and a well-calibrated success probability remain different evidence objects.

An adverse result can reveal a dependency the initial model missed. If several supposedly independent signals fail together, the institution should investigate their shared mechanism before simply requesting another of the same kind. The paper makes no empirical assertion about any actual instrument. It identifies a plausible failure that should have an escalation rule. The evidence map should be revised with the new information, and the decision model should be rerun only when its updated assumptions and scope are sufficiently defined.

Source independence and method diversity are also different. A new analytical method can interrogate the same data and expose a modelling weakness without supplying an independent observation of the state.

That can be valuable for assurance. It should be described as a robustness or method challenge rather than counted as another independent confirmation. Conversely, a new observation can add information even when processed through the same model. The committee needs the evidence object's origin and role, not merely a tally of approaches or publications.

The practical dependency map should connect the current prior, proposed signal, validation evidence and action payoff estimates. A study might share assumptions with the payoffs as well as the state model. If a favourable instrument result is used to justify a larger payoff that already assumed that result, the appraisal can count the same improvement in both probability and consequence. The author should trace the bridge and retain a consistent revised fixture. A narrative statement that inputs are conservative does not replace this coherence check.

A duplicate-evidence walkthrough can expose the dependency before any joint calculation is attempted. In a hypothetical packet, the prior already reflects an engineering assessment and a proposed second report summarises that same assessment through another publisher. A new document has arrived, but the underlying state evidence has not necessarily changed. Mark the shared assessment at both uses and ask what additional observation the second report supplies. If it supplies only interpretation, treat the task as interpretation rather than another independent likelihood update. If it supplies new observations with shared measurement error, a joint account may be required. Counting publishers or methods alone cannot determine which branch applies.

The order of acquisition should be assessed through the decision remaining after the earlier result. A hypothetical first report may make the same feasible action best across all outcomes of a later instrument, leaving little action value for the later commission even if that instrument remains accurate. Another first result may move the institution near an active switch and make the later distinction useful. These conditional branches explain why standalone values should not be added. The single-signal studies illustrate the action-change mechanism, while a numerical package value would require the joint and sequential method absent from the app. The dependency map should identify that missing object rather than supply invented combined probabilities.

Evidence and scope

NIST structures risk assessment through preparation, conduct and maintenance within organisational risk management.

Evidence: Guide for Conducting Risk Assessments

Completed illustrative evidence dependency map

No joint likelihoods are invented. Unresolved relationships remain visible.

Completed illustrative evidence dependency map			
Evidence object	Possible shared input	Question before combining	Current treatment
Current prior	Existing technical screen	Is the new signal already partly known?	Declared pre-signal information set
Instrument likelihoods	Validation population	Does selection alter intended-use performance?	Fixed illustrative assumptions
Second proposed test	Same underlying condition	What are joint outcomes given each state?	Outside single-signal model
Expert judgement	Shared reports or panel	Is agreement independent evidence?	Elicitation provenance required
Payoff matrix	Same favourable operating assumption	Is benefit counted in both likelihood and payoff?	Perimeter coherence review
Repeated measurement	Shared systematic bias	What uncertainty does repetition resolve?	Measurement bridge unresolved

Commission incremental evidence against the current information set

Trace underlying data and assumptions before counting report agreement. Keep a new observation distinct from a new method applied to old data. Do not add standalone signal values or combine marginal likelihoods without an appropriate dependency account. Use a joint or sequential model when those relationships determine the purchase, and preserve the single-signal result as a scoped illustration.

- 1. Map evidence already embedded in the prior.
- 2. Identify shared data, causes and expert inputs.
- 3. Review marginal versus joint performance.
- 4. Reassess incremental value after other evidence arrives.

A qualitative dependency review can be enough for screening

Not every decision needs a large joint model. A qualitative map can identify major overlaps, remove redundant commissions and show that remaining uncertainty is unlikely to change a bounded screening choice. The report should state that limited purpose. When a consequential recommendation relies on the exact benefit of several correlated signals, a qualitative map alone cannot establish the combined value.

No joint or adaptive signal computation is implemented

The engine evaluates a single binary signal with fixed likelihoods. It does not estimate correlations, calculate a joint likelihood table, optimise a test sequence or integrate repeated measurement uncertainty. The dependency worksheet identifies required evidence and method boundaries; it supplies no new quantitative result.

Decision implications

- A report count is different from an independent evidence count.
- Standalone signal values cannot simply be added to value a research package.
- Joint and sequential methods must be implemented before their results are claimed.

GOVERNANCE AND APPLICATION

14

A pilot needs an identification argument

A pilot evaluation design

A pilot finding needs a claim that its design can support

A predictive signal and a causal intervention answer different questions. A signal can help the institution infer which state is likely, while an intervention can change what happens under an action. The binary information model assumes the state probabilities are supplied and shared across the action menu. It does not estimate how Commit, Stage or a pilot changes the probability of success. A report should keep this distinction visible so that a useful posterior is not mistaken for evidence that a proposed programme causes a favourable outcome.

A pilot can produce several evidence objects. It may test a component under specified conditions, measure operating behaviour, explore user response or compare an intervention with a counterfactual. These purposes require different designs and claims. A technically successful trial can support a component-performance statement without establishing

system-wide causation. A before-and-after improvement can be informative for monitoring while remaining insufficient to identify the intervention's effect. The committee should state what the pilot is intended to establish before it uses the results in a payoff or likelihood model.

The app's Stage action is not a causal pilot model. Its success and failure payoffs are supplied aggregate consequences. The engine does not infer those consequences from a trial or estimate the probability of a later successful phase. A reader can use Stage as a hypothetical option to understand the information mechanism, but the name does not confer causal evidence. If a real pilot is expected to alter the state, expand the action set or generate a sequence of findings, its commissioning analysis needs the appropriate causal and sequential structure.

The decision question should determine the identification question. If the committee wants to know whether an operating change improves a specified outcome, the study needs a credible account of what would have happened without that change. If it wants to know whether a component meets a requirement, a bounded test may be appropriate. If it wants to know whether a market signal predicts demand, predictive validation is relevant. Conflating these purposes can lead to a study that is well executed for one question but insufficient for the action used to justify its purchase.

The counterfactual can be difficult to establish in a live industrial setting. Other changes may occur during the trial, participation may be selected, operators may adapt and external conditions may move. The paper does not solve these problems by inventing a controlled field case. It identifies them as design obligations. A committee can use a proportionate evaluation method when its assumptions and limitations are understood, but it should not describe a measured change as a causal effect merely because the programme preceded it.

Selection can affect both prediction and intervention claims. A pilot may operate at the most prepared site, use the strongest team or admit participants most likely to succeed. Its result can be relevant to that bounded deployment while failing to represent a wider population. Transferring it to a larger action requires a bridge covering the conditions that differ. The app will reproduce a supplied likelihood regardless of whether the transfer is defensible. The source and input registers should therefore preserve the trial population, operating conditions and intended deployment scope.

Implementation is part of the causal question when the intervention depends on actual behaviour. A programme described in a design document may differ from what operators receive or use. An evaluation should distinguish assignment, delivery and adoption where those distinctions affect interpretation. A favourable outcome among adopters does not automatically identify the effect of making the programme available to all intended users. The paper keeps these questions conceptual because the binary engine has no participant data, treatment assignment or causal estimation routine.

A pilot can also be valuable without supporting a causal claim. It may reveal an integration constraint, expose an unrepresented failure state or create a new feasible action. Those findings can change the decision architecture. The commissioning record should report that contribution directly and update the relevant state or menu. It should not claim that the original fixed-menu EVSI measured the value of discovering a wholly new design. An exploratory pilot and an instrument used to classify a predeclared binary state have different information mechanisms.

The study should specify the unit of inference. A component, an operator, a facility, a network and an institution can have different mechanisms and dependencies. A result at one unit may not apply at another. The committee should ask whether the evidence supports the consequence level used in the payoff matrix. A component test that reduces uncertainty about a physical parameter can inform a larger model, but that larger model needs its own assumptions and assurance. The transition should be documented rather than concealed behind the phrase pilot validated the project.

There is a similar distinction between model verification and empirical validation. Reproducing the app confirms that the implementation applies its declared formulas to the frozen fixtures. It does not validate the state definition, payoffs or likelihoods against field outcomes.

Reproducing an operating simulator confirms its computation but does not necessarily establish that its behaviour represents the actual system. The decision dossier should state which property each check supports. This avoids accumulating reassuring technical receipts that answer a different question from the committee's evidence need.

The causal evidence register should record the intervention, target outcome, counterfactual logic, relevant population and design assumptions. It should also identify alternative explanations and the limits of generalisation. A reviewer can then judge whether a finding supports the intended action. The register does not impose one universal study

design. It requires the claim to match the design actually used. If the available design supports only an association or a bounded demonstration, that is the finding the report should publish.

An evaluation can influence a later probability model without becoming a universal frequency. The analyst may use a trial to update belief about a specified operating condition, provided the likelihood bridge is supported. It may use an effect estimate to revise supplied action payoffs, provided the consequence model is documented. Those are different paths from evidence to decision. The paper's fixtures do not derive either path from actual trial data. A real dossier should retain the original evidence object and the transformation used to create each input.

A negative pilot finding should be routed through the claim being tested. In a hypothetical trial, the intervention is delivered as designed but the target outcome does not improve; that challenges its effect under the tested conditions. If the intervention was not delivered, the trial informs implementation and does not isolate the intended effect. If the measurement cannot distinguish change from its counterfactual, the effect remains unresolved. If the target population differs materially from intended use, transfer requires a bridge. These branches preserve an adverse finding while preventing it from becoming a universal failure label. They can require a richer signal set than the binary study rather than another reassuring interpretation of the same output.

A completed pilot charter should assign each proposed output its next use. A compatibility demonstration can admit a design to further testing; a supported causal estimate can inform a consequence account within its population; an implementation failure can redesign delivery; discovery of a new option can reopen the action menu. A hypothetical sponsor should not take the first route's successful technical result and claim the second route's payoff improvement. The inference bridge must name the output being used and the condition under which that use is defensible. This instrument preserves the pilot's useful contribution even where it supports a narrower decision than the original sponsor hoped.

Evidence and scope

Magenta methodological guidance distinguishes evaluation approaches and the conditions required to interpret their findings.

Evidence: [Magenta Book Annex A: analytical methods for use within an evaluation](#)

Completed illustrative pilot evidence charter

This is a hypothetical design record; no trial observations or effects are invented.

Completed illustrative pilot evidence charter			
Question	Evidence object	Claim permitted if supported	Claim withheld
Component meets a requirement	Bounded technical test	Performance under stated test conditions	Whole-project success probability
Signal predicts state	Relevant validation study	Conditional predictive performance	Causal effect of acting on signal
Intervention changes outcome	Credible counterfactual evaluation	Effect within design and population	Universal deployment effect
Operators can implement response	Delivery and adoption record	Defined implementation feasibility	Guaranteed realised benefit
New option is discovered	Documented feasible design	Expanded action menu	Value already captured by old EVSI

Match the claim to the study design

Define whether the commission seeks prediction, demonstration, causal effect or option discovery. Record its unit of inference, intended population and alternative explanations. Carry the supported finding through an explicit bridge into the decision inputs. Publish the limited claim the evidence supports, and use a richer model when the intervention changes state probabilities or creates sequential choices.

- 1. Name the evidence purpose before testing.
- 2. Specify counterfactual logic for an effect claim.
- 3. Review implementation and transfer conditions.
- 4. Preserve the evidence-to-input transformation.

Perfect causal identification is often unavailable

An institution may need to act with imperfect evidence. That is a reason to state uncertainty and use a proportionate decision method, not to describe an unsupported causal effect as established. A limited trial can identify constraints, support a bounded choice or justify further investigation. The committee should know which uncertainty remains and what safeguards or revisions respond to it.

No causal or participant-data estimation is implemented

The binary engine has no treatment assignment, counterfactual data, effect estimator, implementation model or generalisation routine. Stage is a supplied aggregate action. The pilot charter identifies the evidence and methodological obligations for a real study without claiming that the worked fixtures establish an intervention effect.

Decision implications

- Prediction, demonstration, causal effect and option discovery are different commissions.
- A named Stage action does not implement a causal pilot model.
- The evidence-to-input bridge should preserve population, conditions and limitations.

GOVERNANCE AND APPLICATION

15 Expert elicitation needs discipline

An elicitation and audit protocol

Record the judgement without pretending it is a frequency

Expert judgement can be necessary when relevant data are incomplete, sparse or poorly transferable. The institution should not disguise that judgement as an observed frequency. It should state the question posed, the information available to the experts, the elicitation process and the way the result enters the decision. A structured judgement can be valuable because its reasoning and uncertainty are reviewable. Its authority does not come from the apparent precision of a consensus number or the prestige of the people who supplied it.

The elicitation question should use the same event definition as the model. Asking whether a project is promising can produce a judgement about attractiveness, feasibility or preference, while the app requires a probability of a defined binary state. Experts may answer different propositions without realising it. A facilitator should therefore test the wording and document the horizon, population, conditions and exclusions. Clarifying the event can be more important than choosing an aggregation technique for numbers that initially meant different things.

Experts should also distinguish their belief about the event from their confidence in the supporting evidence. A specialist may offer a probability while emphasising that the estimate depends heavily on an unresolved operating condition. Another may feel confident about the process but assign a less favourable likelihood. The register should retain both the

event judgement and its evidence limitations. It should not convert a confidence adjective into an unsupported probability or use a narrow display range to imply that disagreement has disappeared.

Selection of experts should follow the knowledge needed for the event. Technical, commercial, operational and institutional uncertainties may require different experience. A large panel can still omit the knowledge of the people who implement the response or bear a consequence. The paper does not prescribe a universal panel size or invent a real panel. It recommends a knowledge coverage record, disclosure of relevant interests and a review of important missing perspectives. The commissioning authority should understand what the panel is equipped to judge and what remains outside its competence.

Information sharing can improve understanding but also reduce the independence of judgements. Experts may begin from the same report, adopt a dominant participant's framing or align after a discussion. The elicitation record should preserve what was shared and when, especially if the final outputs will be treated as separate evidence contributions. A consensus can be useful for a committee decision, but its existence is different from independent confirmation of the state. The dependency chapter's distinction applies equally to a panel and to a stack of published reports.

The facilitator should make reasons for judgement visible. A point estimate without an explanation gives reviewers little ability to assess transfer or identify what new evidence could change it. Experts can describe mechanisms, critical assumptions, conflicting observations and conditional scenarios. These explanations can improve the action architecture even when the numerical range remains broad. The purpose is to capture decision-relevant knowledge rather than extract a convenient percentage that the optimiser can process without further discussion.

The process should reduce avoidable distortions without claiming to eliminate judgement. Participants may be influenced by an initial number, a desired project outcome or incentives associated with a recommendation. A structured protocol can separate initial estimates, discussion and revision, and can require explicit reasons for changes. The institution should record the protocol and any departures. The paper does not claim that a particular protocol guarantees calibrated probabilities. Its practical value is to make the conditions of judgement sufficiently transparent for an independent reviewer to challenge.

Aggregation is a modelling choice. Averaging individual estimates can be simple and reproducible, but it does not automatically resolve different event definitions, expertise, dependencies or interests. Another aggregation rule may be appropriate in a defined context, provided its rationale is recorded. The binary app does not combine panel judgements or infer a distribution from them. A real elicitation output should be processed through the declared method before it becomes a supplied prior or likelihood, and the transformation should be retained with the evidence.

Ranges should have a meaning. A lowest and highest estimate across experts is different from a probability interval supplied by one expert, and both differ from an interval estimated from data. The committee should not treat them interchangeably. If the elicitation is used only to select plausible sensitivity fixtures, it should say so. If it supports a probabilistic model of parameter uncertainty, that model requires an explicit method. The paper's contrasting studies are authored scenarios, not a panel distribution or an empirical confidence interval.

Calibration and performance evidence require a relevant record. A specialist's experience can justify including them in a panel without proving the numerical accuracy of every judgement. If a commission relies on a claim of expert calibration, it should identify the questions, outcomes, scoring method and transfer conditions that support it. The paper makes no such claim for hypothetical role holders. It keeps the elicitation record at the level it can honestly support: a structured process for documenting judgement and identifying the evidence needed to assess its numerical use.

The institution should preserve disagreement that changes the decision. A report that publishes only a consensus prior can conceal that credible alternatives select different actions. The app can compare explicitly supplied fixtures to show that sensitivity, while the source register retains the reasons for each alternative. The committee can then decide whether to commission evidence, choose a more robust action or accept a documented exposure. Suppressing consequential disagreement to make a decision look settled defeats the purpose of eliciting informed judgement.

Judgement may need updating when new evidence arrives. The update should have a stated trigger and a preserved history. An expert's revised number can reflect a new observation, a corrected event definition or a change in incentives and interpretation. The reasons matter. A new number should not silently overwrite the original basis of the research

purchase. The institution should retain enough context to determine whether the current prior actually represents the information set it claims to represent.

Disagreement can be diagnosed before averaging expert numbers. Hypothetical experts may be answering different success events, using different information sets or applying different assumptions about implementation. Resolve those differences into a common question or retain separate scenario propositions. If they share the question but disagree about likelihood, preserve the reasons and test the decision under the distinct judgements before selecting an aggregation method. If every defensible judgement produces the same available action, further convergence may have little immediate decision use. If the policy changes, the disagreement itself becomes part of the commission's sensitivity. This routes elicitation effort to a consequential distinction rather than make consensus a production target.

A protocol should retain the judgement that discussion displaced. In a hypothetical panel, an expert revises a view after learning of a previously missing observation. Record that observation and the changed rationale. A revision after hearing the chair's preferred action has a different interpretation and should not be described as new empirical support. The retained individual and revised accounts allow the facilitator to distinguish learning, clarified definitions and social convergence. They do not establish calibration without a relevant performance record. This makes the elicitation output inspectable and keeps a formatted panel conclusion from acquiring the status of an observed frequency merely because participants finally agree.

Evidence and scope

EFSA describes formal elicitation as a way to manage psychological biases and aggregation, with material resource demands.

Evidence: [EFSA Tutorial on Uncertainty: Overview of methods for Uncertainty Analysis](#)

Completed illustrative expert-judgement protocol

Roles and procedures are hypothetical. No elicited probability or actual participant is asserted.

Completed illustrative expert-judgement protocol		
Stage	Specified record	Challenge or acceptance test
Question definition	Binary event, horizon and conditions	All participants judge the same proposition
Knowledge coverage	Technical, operating and consequence roles	Material expertise gaps are visible
Initial judgement	Independent reasoning and stated range	Origin and uncertainty are retained
Shared discussion	Evidence and disagreements recorded	Changes are explained, not compelled
Aggregation	Declared transformation into model input	Dependencies and differing meanings reviewed
Decision sensitivity	Explicit alternative fixtures	Consequential disagreement remains visible
Update	New evidence trigger and version history	Original commission basis remains recoverable

Use judgement with a preserved question and process

Specify the event before collecting estimates, record knowledge coverage and interests, and preserve reasons for disagreement. Distinguish individual ranges from data intervals or a joint probability model. Declare any aggregation method outside the binary engine. Let decision sensitivity determine whether a more formal process is necessary, and retain the evidence that would warrant an update.

1. Align event definitions across participants.
2. Preserve individual reasons and shared information.
3. Declare aggregation and range meanings.
4. Test whether disagreement changes the action.

Formal elicitation can become disproportionate

A detailed protocol can consume resources without changing a bounded decision. The commissioning team should first identify the uncertainty that could alter the action and assess available data. A concise documented judgement can be appropriate for a screen, provided its status and limitations are explicit. Formality is justified by the decision need and risk of misinterpretation, not by an aspiration to make every input look statistically certified.

The app does not elicit or aggregate judgements

The engine accepts fixed prior and likelihood inputs. It has no expert-scoring, bias-correction, aggregation or parameter-distribution method. EFSA material is used in its food-safety assessment context for narrow process guidance; it supplies no numerical probability for the illustrative commercial decision. The protocol identifies an additional evidence workflow, not a hidden model feature.

Decision implications

- Experts need a shared event definition and an explicit information set.
- Disagreement and range meanings should survive aggregation.
- The binary engine accepts judgement inputs and does not validate the elicitation process.

GOVERNANCE AND APPLICATION

16

Stopping research can be rational

A sequential evidence policy

Stop or redesign when the remaining decision no longer justifies research

Stopping research is a decision about the remaining commission, not a verdict that the original team was foolish. New evidence, a changed opportunity or an unavailable response can make continued work unattractive even when the initial purchase was reasonable. The review should compare the remaining benefits and avoidable costs using the current action architecture. It should preserve the original charter so that a revision is understandable, while allowing the institution to respond to conditions that were not known when the study began.

The relevant cost at a continuation review is the cost that can still be avoided. A fee already irreversibly paid is not recovered by stopping. A remaining field trial, analytical stage or operating disruption may still be avoidable. A contract can also impose a termination consequence that changes the comparison. The paper does not calculate a particular contract or invent termination terms. It recommends a remaining-cost register that distinguishes sunk commitments from future choices and identifies the evidence needed to value each component.

The current decision can change before the study finishes. A new external result may make the proposed signal redundant, a revised design may move the action boundary, or an approval restriction may remove the response that generated most of the original value. The continuation review should update the prior, menu and payoff perimeter only when the

reason is documented. A team should not keep quoting the original EVSI as a justification for every additional research stage after its assumptions have ceased to apply.

The no-policy-change and restricted-action fixtures identify distinct reasons to reconsider a commission. In the former, the current prior is already favourable enough that both branches choose the same action. In the latter, the high-upside response is unavailable. Neither case is a dynamic stopping calculation. They are static comparisons that show how a current decision can have less information value under changed conditions. A real continuation rule would need the timing, remaining evidence process and cost of stopping or continuing to be represented explicitly.

The study may also cease to be feasible. The intended population may be inaccessible, an instrument may not perform in the required conditions, or the remaining window may be too short for credible analysis. Continuing can produce a deliverable that looks complete while no longer answers the commissioning question. A method checkpoint should therefore review the evidence object's relevance and quality as well as progress against a schedule. Completion of planned tasks is not sufficient when the plan has lost its connection to an actionable decision.

A partial finding should not automatically become a valid early signal. The team may have observed a subset of the intended data or an ambiguous result that does not follow the commissioned likelihood model. Acting on that partial result can require a new analysis. The paper's binary engine does not implement sequential statistical monitoring or correct for repeated inspection of accumulating data. A governance decision to review the commission is different from a methodologically valid stopping rule for an experiment. The record should keep these questions separate.

The committee should predeclare material review triggers where possible. These can include a changed action deadline, a removed response, a new evidence source, a failure of study access or a material cost revision. A trigger prompts a review; it does not determine the result automatically. The institution may continue because a broader declared purpose remains worthwhile, redesign because another instrument better fits the current decision, or stop because the remaining value no longer justifies the effort. The reasons should be recorded in the same decision language as the original charter.

Stopping can preserve resources while creating obligations. A partially completed study may require safe closure, disclosure of material findings, data preservation or an operating handback. These conditions belong in an appropriate institutional review. The analyst should not assume stopping is costless merely because future analytical work is cancelled. Nor should it invent a general legal duty from a methodological framework. The relevant owners should identify the obligations that actually apply and include their consequences or constraints in the continuation decision.

The decision should avoid escalation of commitment. An institution can be tempted to finish a study because much has already been spent, because stopping might appear embarrassing or because a sponsor promised a report. These incentives can be acknowledged without making claims about actual people. The practical control is to ask an independent reviewer to assess remaining value and cost against the current decision. Preserving the initial rationale helps learning; it should not bind the team to complete work whose commissioning purpose has materially disappeared.

The opposite error is to stop because a preliminary finding is uncomfortable. A sponsor may prefer to end a study that begins to challenge the desired commitment. A predeclared charter and review path can make that behaviour visible. The review should distinguish lost usefulness from an adverse but relevant result. Evidence that makes the committee choose Hold may be exactly the response that justified the original commission. Disliking that outcome is not a methodological reason to suppress the finding or redefine the study's purpose.

The continuation decision should retain available alternatives. It may be possible to narrow scope to the uncertainty that still controls the action, use already obtained evidence, switch instruments or redirect the study to a later declared decision. Each alternative needs its own cost, feasibility and claim scope. A report cannot simply call an incomplete original study a complete new one. The revised deliverable should say what changed and which findings are still supported. The institution should maintain the distinction between reusing valid evidence and retroactively inventing an objective.

A positive gross information value is not sufficient to continue at any price. The expensive-signal fixture shows the basic acquisition logic: a signal can improve the policy while its cost exceeds that improvement. At a continuation review, the relevant price is the supported remaining avoidable cost under the current perimeter. The app can compare a

supplied revised aggregate cost in a new explicit fixture, but it does not determine that cost or a sequential continuation strategy. The record should identify that distinction before relying on the calculation.

The completion disposition should follow what remains usable. A hypothetical commission stopped after loss of data access may leave a valid specification and no accepted signal. A commission narrowed after a changed action window may leave a useful technical finding for a later decision but no value claim for the original response. A commission finished with an adverse accepted signal can fulfil its purpose by supporting the predeclared withdrawal or hold. Preserve each evidence object with its actual status and withdraw only the inference it no longer supports. This prevents stopped from becoming an empirical failure state and prevents uncomfortable from becoming a reason to erase valid evidence.

A continuation review needs paired reasons for continuing and stopping. In the hypothetical no-policy-change branch, ask whether another legitimate purpose remains and what it would cost to fulfil. In the restricted-action branch, ask whether restoring the response is a feasible separate intervention, rather than assume the research will create authority. In the expensive-signal branch, compare the remaining avoidable burden with the decision still available, leaving sunk spending outside the argument for continuation. These are governance dispositions; they supply no adaptive experimental stopping rule. Their practical contribution is to make both a stop and a continuation answerable to the current purpose, instead of favouring whichever result preserves the original commission.

Evidence and scope

GAO cost guidance links technical baselines, documented assumptions, sensitivity and updates with actual information.

Evidence: [Cost Estimating and Assessment Guide](#)

Illustrative calculation. Assumptions and sensitivity are published with the model.

Illustrative mechanisms that can defeat an information purchase

These are separate static fixtures. They are not an optimised sequential stopping policy.

Illustrative mechanisms that can defeat an information purchase			
Mechanism	Gross signal value, USDk	Supplied cost, USDk	Net value, USDk
Current policy does not change	0.00	5.00	-5.00
A major response is unavailable	3.45	5.00	-1.55
Acquisition cost is excessive	33.20	80.00	-46.80

Completed illustrative continuation review

No actual sunk payment, contract or study result is asserted.

Completed illustrative continuation review		
Review field	Current question	Recorded outcome requirement
Decision	Can the institution still change a relevant action?	Current menu and authority accepted
Evidence	Will the remaining study answer the stated question?	Method and use context remain valid
Cost	Which future costs are avoidable?	Sunk and remaining components separated
Timing	Can the result be reviewed before the action closes?	Usable window and fallback recorded
Obligations	What follows from ending the commission?	Appropriate owner reviews actual duties
Alternatives	Can scope, instrument or purpose be revised honestly?	New deliverable and claims specified
Challenge	Is the decision driven by sunk spending or an adverse result?	Independent reasons preserved

Review the remaining commission on its current merits

Use documented triggers to reassess decision relevance, instrument feasibility, remaining avoidable cost and usable timing. Preserve the original charter and valid evidence, but allow a justified stop or redesign. Separate governance continuation review from experimental stopping methods. Record obligations and the exact claim scope of any revised deliverable.

1. Identify remaining benefit and avoidable cost.
2. Check the response menu and action window.
3. Distinguish adverse evidence from lost study relevance.
4. Preserve the revision and withdrawal record.

Frequent reviews can disrupt a coherent study

A study needs stable methods and enough continuity to produce interpretable findings. Governance checkpoints should be proportionate and should not invite arbitrary redesign after each inconvenient observation. Predeclare material triggers, preserve methodological rules and use independent review when a change is consequential. Flexibility is valuable when it protects the decision purpose without destroying the evidence design.

No adaptive stopping or termination-cost engine is present

The app evaluates a fixed binary signal and aggregate price. It does not optimise a sequence, calculate remaining contractual costs or provide statistical stopping rules. The worksheet is an original governance framework, while the bound table illustrates static purchase failures. A real continuation calculation requires the additional evidence and method appropriate to its timing and obligations.

Decision implications

- Continuation depends on remaining avoidable costs and the current decision.
- An adverse but relevant result can fulfil the original commissioning purpose.
- Governance review and experimental stopping rules are different methods.

GOVERNANCE AND APPLICATION

17 A committee must own its loss function

A consequence and assurance ownership register

Somebody must own the consequence account

The loss function should have an accountable owner. An analyst can calculate expected money accurately while using consequences that no authorised institution has accepted. A sponsor can provide optimistic returns, an operator can identify service burdens, and a finance team can

supply cost estimates, but someone must approve the perimeter that combines them. This approval is a judgement about what the decision represents. It should be visible before the research purchase is presented as an institutional recommendation.

Ownership does not mean one person invents every input. It means the decision record identifies who supplies evidence, who checks meaning, who accepts the objective and who authorises the response. These roles can be separate. An independent reviewer can challenge a method without assuming the committee's mandate. An operating specialist can reject a feasibility premise without determining the institution's risk preference. The charter should preserve those distinctions so that an expert signature beside a technical parameter is not misread as approval of the entire commitment.

The payoff perimeter should identify incidence explicitly. A saving to one department can be a cost to another; a sponsor's upside can coexist with an operator's exposure. A single institutional payoff may aggregate these consequences if the aggregation is justified and accepted, but the underlying allocation should remain visible. If the commission affects interests outside that institution, the decision owner should know which review process applies. The simple app does not infer affected parties or calculate distributional fairness. Those questions belong in the declared governance and consequence account.

The commissioning budget also has an owner. The person buying evidence may not be the person who captures its benefits, and that difference can distort the purchase discussion. A research team may be asked to justify a study by counting broad organisational benefits while the response department has no budget or incentive to implement the result. The commission should identify both the funding owner and the action owner. If a shared benefit is real, the institution can organise a coherent budget and response path rather than pretending the binary model resolves an internal allocation dispute.

The institution should ask who can dispute a cell. A failure consequence may depend on operating knowledge that the investment committee does not possess. An upside estimate may rely on a commercial assumption that the technical team cannot certify. A protected condition may require an authority outside either group. The input register should route each challenge to the right owner and preserve unresolved disputes. A forced consensus can make a matrix easy to process while concealing the exact assumptions most likely to defeat the recommendation.

The approved objective should be versioned. If the committee decides to widen the consequence perimeter, add a constraint or change the admissible menu, the numerical comparison should be rerun with a documented reason. The report should not overwrite the old model and present the revised optimum as though it had always supported the original commission. Version history enables reviewers to understand whether the institution learned, changed priorities or corrected an error. It also makes a later disagreement more specific than a vague claim that the numbers moved.

Ownership extends to the evidence instrument. The supplier should describe what the study can establish and which assumptions it uses, while the client should confirm the intended decision and use conditions. A mismatch can occur when a technically appropriate study answers a different question from the client's payoff matrix. The acceptance process should compare the delivered evidence object with the commission, then review the bridge into inputs. A polished report is not sufficient when its findings cannot support the likelihood or consequence it was purchased to inform.

The model owner should be able to reproduce the numerical object independently of the narrative author. The frozen input, output and binding files make that possible in the public examples. In a real institution, the owner should retain the exact computational version, input provenance and approval history. A result copied from a presentation without its fixture cannot support a reliable later challenge. The role is to preserve the calculation's identity and limits, not to claim that reproduction alone validates the empirical assumptions or the institution's decision.

The challenge owner should have room to disagree. A review that checks only whether the report is internally consistent can miss an unacceptable objective or omitted alternative. The adversarial question is whether an admissible change in evidence, consequence or action set overturns the proposed commission. The paper's original eight contrasting studies provide several such counterexamples. A useful challenge record should identify the premise, the contrary fixture or evidence, the result and the disposition. It should not treat every objection as a wording issue that can be closed without changing the decision.

The organisation also needs an escalation path for unresolved conflicts. If the technical team accepts a signal but the action owner cannot implement the response, another accuracy discussion will not solve the gap. If the committee accepts a monetary optimum but an affected interest identifies a binding condition, a revised font or disclaimer is

irrelevant. The decision record should identify the authority that can resolve the conflict, the evidence needed and the interim action. This makes the report useful without claiming that the analyst owns every institutional judgement.

Conflicts of interest should be managed through disclosure and appropriate review. A study provider may benefit from a larger commission, a sponsor may benefit from commitment, and an operating team may prefer a familiar process. These are plausible role incentives, not accusations about actual people. The record should make the interests relevant to a parameter or decision visible and avoid giving a conflicted owner unchecked authority over the entire chain. The appropriate institutional process should determine the controls; the binary model cannot detect incentives from its numerical fields.

The burden of review should be proportional to the consequence and sensitivity. A limited screen may need a concise owner register and a transparent assumption set. A consequential commitment with disputed payoffs and concentrated exposure needs a stronger review path. The committee should justify the chosen assurance level and identify what remains unresolved. More signatures do not automatically mean better assurance when every signer checks the same narrow property. Coverage of the consequential assumptions matters more than a ceremonial list of approvals.

A correction can propagate through different objects. A hypothetical transcription error in a likelihood changes the fixture, posterior and branch policy and requires those outputs to be rerun. A newly adopted protected condition can change the admissible menu while leaving the old likelihood record intact. A revised judgement about a payoff changes the consequence account and its authority record. In each branch, identify the affected recommendations and recipients before withdrawing or replacing them through the authorised process. This separates an error correction from a changed mandate and prevents a corrected headline from circulating alongside an unchanged response instruction whose premises no longer hold.

Ownership can be tested through a disagreement drill. A hypothetical Operations owner disputes the payoff assigned to a failed deployment, while the Instrument lead accepts the likelihood account. The model owner can reproduce both accounts but cannot settle the institutional consequence merely by pointing to correct arithmetic. Route that cell to its authorised consequence owner, preserve the competing rationale and identify the feasible interim action. If the dispute concerns the laboratory population instead, route it to instrument assurance. A responsibility

register works when it moves each disagreement to an actor able to decide the actual premise, rather than send every objection back to the numerical analyst.

Evidence and scope

DOE relates verification effort to performance risk, savings magnitude and project complexity.

Evidence: Measurement and Verification for Federal Energy Savings Performance Contracts

Completed illustrative consequence and assurance ownership register

Functional roles are hypothetical; the institution must assign actual authority before use.

Completed illustrative consequence and assurance ownership register			
Object	Supplier or author	Accountable acceptance	Independent challenge
Event and prior	Evidence analyst	Decision owner	Method reviewer
Instrument likelihoods	Study provider	Evidence acceptance lead	Validation specialist
Payoff cells	Commercial and operating owners	Committee objective owner	Consequence reviewer
Computational fixture	Model maintainer	Analytical assurance owner	Reproduction reviewer
Action admissibility	Response owner	Delegated authority	Mandate reviewer
Research budget	Commissioning sponsor	Budget authority	Purchase-value challenger
Correction	Evidence and model owners	Publication or decision authority	Independent disposition reviewer

Assign ownership to the entire chain

Name the owners of evidence, payoff incidence, objective, computation, response and challenge. Preserve disputes and version changes. Route a challenge to the authority that can resolve it, rather than treating every objection as a model defect. Require correction ownership before using the result for an institutional commitment.

1. Assign a supplier and acceptor for each material premise.
2. Identify consequence incidence and unresolved disputes.
3. Separate reproduction from empirical and mandate acceptance.
4. Maintain an escalation and correction path.

Extensive governance can slow a bounded decision

A concise register can be enough when the consequence is limited and assumptions are not contentious. The answer is proportionate coverage, not automatic complexity. The institution should still know who owns the objective and response. Removing those roles for speed can create a study that arrives with no accountable path from its findings to an action.

The ownership framework does not establish authority

The worksheet is an original governance proposal for adaptation. It does not identify actual delegated powers, resolve affected-party rights or certify institutional controls. The app contains no governance engine. Appropriate institutional evidence and review must establish the owners and permissions relevant to a real decision.

Decision implications

- Evidence, objective, computation and response can have different owners.
- A challenge needs an authority capable of resolving its premise.
- Correction ownership belongs in the original decision architecture.

GOVERNANCE AND APPLICATION

18 Evidence quality must survive publication

A claims and correction control framework

Make the evidence object inspectable without overstating it

Publication should expose the decision argument within its evidence authority. Identify public primary guidance, original analysis, illustrative input assumptions and canonical outputs where each claim appears. A primary methodological statement does not validate commercial payoffs; arithmetic parity does not demonstrate field performance. Keeping these distinctions beside the result lets a reader inspect its basis without inheriting an unsupported commercial or empirical conclusion.

The canonical manuscript serves as the common narrative for the PDF and chapter reader. This avoids maintaining separate summaries whose claims drift from the worked study. A quantitative block refers to named ledger tokens, which in turn bind to a specific model fixture or source-supported claim. The renderer resolves those tokens into the displayed values. The method gives reviewers a concrete route from sentence or table cell to its basis. It does not replace semantic review of whether the basis supports the interpretation.

The frozen numerical object consists of an explicit input fixture, its canonical engine output and scalar bindings. The output is retained as full structured data, rather than only the displayed rounded cells. A reader can reproduce the app computation and compare it with the frozen object. Bindings identify the values selected for publication and their pointers. This is especially useful when a chart or table shows only part of a large output. Reproducibility depends on preserving the selected computation and the meaning of its inputs together.

Keep the illustrative label, title, basis, units and limitation attached to each quantitative exhibit. A copied table or independently read chapter may lose the paper-wide note. The local disclosure must therefore prevent the figures being mistaken for observed industry rates, actual client cases or quoted contracts. These examples contain no private or empirical records from which such an interpretation could be supported.

Source cards should be narrow and locator-specific. Their purpose is to show what public material was actually retrieved, what limited claim it can support and what it cannot establish. A source title alone does not prove that a report's interpretation is justified. The record should retain the publisher, edition, exact URL, retrieval object and relevant location. If only a publisher description was readable, the card should say so. An inaccessible full document should not be represented as read merely because another page cites its title.

The shared catalogue holds one public source record for reuse across papers. References retain its scope and limitation; narrow local notes connect it to the decision without repeating long guidance paraphrases. Original mechanism and fixture reasoning must remain independent analysis. Keep a sourced assertion sourced even when its word budget or citation obligation is inconvenient: the claim record follows its substance, not the label that reduces production work.

Rounding should be declared and should not change the stored result. The paper displays selected monetary values at a consistent precision and retains raw outputs in the replication files. A probability stored as a fraction can be displayed as a percentage when the binding format and units make that conversion explicit. This prevents a display convention from silently becoming a new computational input. A reviewer should be able to understand a rounding difference without needing to infer whether the model, table or narrative used a different fixture.

Figures need their own reproduction object. A quantitative chart can have correct-looking axes while displaying no actual series, a clipped point or a value that lacks a reciprocal ledger binding. The plot specification should

identify the selected output pointers, drawing ranges and labels. The shared renderer recreates the SVG and compares exact bytes, rejecting empty series and invalid coordinates. This checks the graphical computation. It does not establish that the author chose a useful comparison or wrote an accurate caption; those remain independent review questions.

Reopen the PDF: declared pages and successful rendering can conceal thin, duplicated or clipped content. The audit identifies chapter boundaries, counts qualifying body pages, checks text geometry and fonts, and renders every page for visual review. Its receipt must name the checked properties and remaining semantic or visual judgements. A mechanical pass leaves a review candidate; independent acceptance requires review of the substantive artifact.

The reader should have access to limitations with the result rather than only in an appendix. The app calculates a fixed binary risk-neutral decision with supplied likelihoods and aggregate payoffs. It does not implement causal estimation, utility elicitation, joint signals, adaptive stopping or a dated delay model. A chapter discussing those methods should state their conceptual status. Publishing a diagram of a broader framework does not make the app calculate it. The paper's claim scope should match the actual implemented methods, not the ambition of its contents list.

Publication should preserve failed retrieval and correction evidence privately where appropriate. A source archive can record a blocked URL alongside a successful publisher response, without presenting the failure as verification. A challenge log can retain withdrawn interpretations and their disposition. The public card should describe the current narrow support honestly. The existence of a failure log is not a defect by itself; it can show that the team distinguished absence of evidence from successful access. What matters is that unsupported content is not silently certified.

The final edition needs a stable identity. The manuscript, sources, claim ledger, computation, PDF, chapter fragments and relevant proof receipts should have consistent hashes. A change to a numerical input requires recomputation and a new binding review. A change to the renderer can require a fresh artifact audit even when the prose is unchanged. A change to the source interpretation can require a new semantic review. These distinctions allow an institution to know which properties of an earlier review still apply to the revised edition.

Record mechanical pass, author proof review, independent methodological review and public delivery as separate states of the frozen edition. Model reproduction or a page threshold closes only its checked property. Independent reviewers still need the full artifact and claims; pending dimensions remain pending until their evidence exists. Updating status on that basis prevents a production milestone from being published as certification or substantive endorsement.

An edition drill should follow a hypothetical corrected signal interpretation across every public object. The raw output can remain correct while a chart legend has reversed the branch labels; repair the legend and its explanatory claim, then verify the displayed response against the retained branch account. If the raw likelihood input was wrong instead, replacing only the PDF leaves the chapter reader and exported response policy stale. Trace fixture, output, binding, figure and displayed conclusion to the same edition identity. This drill turns publication integrity into a specific object-consistency test while preserving the separate question of whether the retained primary evidence supports the premise at all.

Evidence and scope

NASA links decision context, alternatives, assumptions, uncertainty and recorded rationale within systems engineering.

Evidence: [NASA Systems Engineering Handbook](#)

Public evidence and reproduction object register

Each object supports a specific check; none establishes all dimensions of acceptance.

Public evidence and reproduction object register			
Object	Reader can inspect	Check supported	Judgement still required
Canonical chapter	Narrative, exhibits and limits	PDF and reader use the same blocks	Argument is substantive and useful
Source card	Publisher, locator and retained evidence	Retrieval identity and scope	Locator supports the cited sentence
Claim ledger	Value, basis and reciprocal block	Selected claim is traceable	Interpretation is warranted
Fixture and output	Full input and computed result	Canonical reproduction	Inputs fit a real decision
Plot specification	Output pointers, labels and range	Exact graphic redraw	Comparison and caption are useful
PDF proof receipt	Actual pages, text and renders	Artifact properties checked	Independent all-page visual acceptance
Challenge record	Objection and disposition	Revision is recoverable	Independent closure is justified

Publish the claim with its basis and boundary

Bind quantities to the frozen computational object, retain raw values and declare display units. Keep source notes narrow and preserve actual retrieval scope. Reopen and render the artifact rather than trusting declarations. Distinguish mechanical pass from independent acceptance, and rebind the edition when its relevant objects change.

1. Connect a displayed claim to its source or fixture.
2. Retain rounding and units in the binding.
3. Inspect every actual rendered page.
4. State the precise review status of the frozen edition.

Too much evidence machinery can overwhelm a reader

A publication can provide a clear main argument and concise exhibits while placing full reproduction objects in a linked desk. The solution is layered access, not removing the objects or forcing every reader to inspect raw JSON. The main result should remain understandable, with enough basis and limits at the point of use to prevent misinterpretation. A motivated reviewer should then be able to reach the exact evidence.

Traceability does not establish semantic support

Hashes, reciprocal bindings and exact redraws demonstrate identity and reproducibility properties. They do not prove that a source supports every interpretation, that a payoff is empirically appropriate or that a framework is useful. Independent domain, adversarial and visual review remains necessary. The production audit is designed to expose those boundaries rather than manufacture acceptance.

Decision implications

- A source locator, raw fixture and displayed claim should be connected.
- Rounding and chart reproduction are artifact properties with explicit limits.
- Mechanical pass remains distinct from independent acceptance.

GOVERNANCE AND APPLICATION

19

Where expected information value is insufficient

A non-monetary constraint framework

The decision architecture must be able to refuse a positive value

An information-value model cannot make every consequence compensable. An institution may face a protected condition, a mandatory investigation or an obligation that determines which actions are allowed. These boundaries should be reviewed by the relevant authority using the actual context. A positive net monetary value for a study or action does not supply permission to breach them. The analytical task is to identify the boundary and carry it into the action architecture, rather than insert an arbitrary penalty that the optimiser can outweigh with sufficient upside.

Use monetary consequences and admissibility conditions to frame the committee's review, without a transaction-specific legal conclusion. The institution must establish applicable rules, rights, duties, interpreting authority and required evidence. The binary app has no jurisdiction or mandate logic: well-formed supplied actions remain conditional, with no implied legal, regulatory or affected-party acceptance. Numerical validity cannot close the institution's review of its operative menu.

A mandatory research purpose can coexist with an unfavourable immediate information purchase result. An institution may need to investigate a condition because of an actual applicable duty, assurance requirement or accepted governance policy. The commission should name that purpose and its authority. The risk-neutral EVSI can still describe a limited monetary use, but it should not be used to cancel the mandatory purpose or claim to quantify its entire value. Mixing the purposes creates confusion both when the result is positive and when it is negative.

The same distinction applies to public evidence. A study can inform accountability, public understanding or a future collective decision that the immediate payoff matrix does not represent. Those benefits need a suitable objective and review process. The paper's current-decision model should not be expanded by a vague narrative claim that transparency is always worth any cost. The institution should specify the audience, evidence need, decision horizon and expected use. A legitimate broader purpose becomes more defensible when it is stated directly rather than attached as an unmeasured rescue term to an unattractive calculation.

Unknown states can create another boundary. The engine represents success and failure as exhaustive categories. A novel hazard, an unanticipated interface or a consequence outside that definition may not fit either cell. Treating failure as a catch-all can be useful only when the aggregate consequence genuinely covers the omitted mechanisms. If it does not, the model is incomplete for the decision. A committee can use scenario stress, a robust constraint or an exploratory investigation while recording that probability support is absent. It should not enter a convenient number to make the uncertainty appear solved.

An instrument can also be inappropriate even when it has positive calculated value. The study method may be unavailable, unacceptable in its intended setting or unable to protect sensitive information. These conditions require evidence and appropriate review. The paper publishes only public sources and illustrative fixtures, so it does not face the private-data assessment of a real commission. Its replication objects should not be treated as a template for publicly releasing confidential operating records or contracts. Transparency should be organised within the actual permissions and legitimate evidence needs.

The institution should distinguish uncertainty about a constraint from a disputed preference. A committee may not know whether an action is allowed; that calls for an authority and evidence review. It may know the action is allowed but disagree about its desirability; that calls for a preference or consequence discussion. It may know both but lack capacity; that calls for implementation and resource analysis. The expected monetary optimum does not resolve any of these by itself. A clear issue register routes the question to the method and owner capable of answering it.

For a safety-related choice, document the relevant consequence, authority and evidence standard, then exclude responses outside the authorised conservative boundary from monetary optimisation. A commission intended to resolve that constraint needs an explicit purpose and

acceptance condition. This proposes no universal threshold or adequate test for an actual safety decision; the commercial likelihood demonstration supplies no safety-critical instrument validation.

The model's failure state should therefore receive a scope review. What types of consequence are included in the supplied aggregate? Which remain uncertain? Which are handled by a separate constraint? Which parties can contest the perimeter? These questions are more useful than a broad warning that all models have limitations. They tell the committee exactly what could make the current calculation unsuitable. The completed boundary register should have named functional owners and clear unresolved dispositions, not only a general disclaimer at the end of the report.

An institution can keep the model as a reference even when a constraint determines the choice. The expected-value result can show the monetary consequence of the accepted restriction, provided its inputs remain suitable. The report should state that the final action follows the admissibility boundary or broader objective, not pretend that the original risk-neutral optimum approved it. This preserves analytical honesty and helps future reviewers understand why the institution declined an apparently attractive response. Numerical transparency is compatible with a decision that rests on other legitimate conditions.

Evidence can reveal that an initially assumed constraint does not apply or that a new constraint does. Either event changes the operative action architecture and should be documented. The restricted-action fixture demonstrates how a removed response changes information value, but it does not estimate the likelihood of an approval or supply the process for obtaining one. A real commission addressing that issue would need a relevant evidence object and authority. The analyst should not borrow the fixture's monetary difference as a universal price for governance or permission.

At publication, preserve a consequence-boundary disagreement even if the action owner accepts the payoff matrix. A relevant reviewer's unresolved objection can prevent acceptance. The institution must resolve the operative perimeter through its review process; reproducing the supplied calculation cannot establish that this is the perimeter it may use. Keeping the objection and its decision effect open makes the fail-closed boundary operational.

A positive result can be refused for different reasons with different consequences for research. In a hypothetical protected-service dispute, changing the monetary payoff does not remove the overriding condition;

the action must remain outside the admissible menu unless the proper authority establishes otherwise. In an incomplete state model, a broader investigation may repair the question itself. In an unavailable implementation route, capability work may precede the evidence purchase. Preserve these branches rather than write one generic refusal beside the positive number. The original calculation remains a scoped reference, while the disposition identifies which premise failed and whether the next task is authority review, discovery or operating preparation.

A mandatory commission needs an independent completion test. If hypothetical research is required to document a protected condition, its acceptance should depend on the specified evidence object and applicable review, even where the immediate binary purchase value is unfavourable. Conversely, a favourable purchase value does not show that the mandatory evidentiary requirement has been met. Keep the monetary reference and the mandate record side by side, with their different purposes intact. This prevents an optimiser from cancelling required work and prevents the institution from presenting mandatory expenditure as a demonstrated discretionary return. No actual legal duty or safety approval is established by this hypothetical disposition.

Evidence and scope

UK multicriteria guidance addresses explicitly structured comparisons when monetary appraisal does not represent every criterion.

Evidence: [Multi-criteria analysis: a manual](#)

Completed illustrative decision-boundary disposition register

This identifies questions and owners. It asserts no actual legal or safety conclusion.

Completed illustrative decision-boundary disposition register		
Boundary question	Required owner or method	Interim disposition
Action permitted?	Appropriate mandate or legal authority	Exclude unresolved action from operative menu
Protected condition represented?	Affected-interest and consequence review	Do not invent a compensating penalty
Mandatory evidence purpose?	Authority for actual applicable requirement	Assess scope on its declared basis
Unknown state consequential?	Method and domain reviewers	Use bounded screen or broader state model
Instrument acceptable for use?	Relevant evidence and operating authority	Withhold unsupported deployment claim
Public release permitted?	Evidence and publication owners	Use authorised public material only
Capacity to respond?	Implementation and resource owners	Choose currently feasible response

Allow an explicit boundary to refuse the number

Identify constraints, protected conditions, mandatory purposes and unrepresented consequences before accepting a monetary optimum as a decision basis. Route each question to its actual authority or method. Keep a scoped reference calculation where useful, but state when the final action follows a different legitimate condition. Preserve unresolved disputes rather than certifying them through reproducibility.

1. Separate admissibility from preference.
2. Review the failure-state perimeter.
3. Name authority for any mandatory commission.
4. Withhold acceptance where consequential support is unresolved.

Constraint language can be used to avoid an inconvenient result

A sponsor can misuse an undefined constraint to ignore evidence. Require the condition, owner, basis and effect on the action to be recorded before or through an accountable revision. A legitimate boundary should be reviewable and applied consistently, including when it obstructs a favoured commitment. The framework protects both against numerical overreach and against arbitrary refusal to learn.

No permission, rights or safety certification is calculated

The app does not determine applicable duties, evaluate affected-party rights, certify a safety case or authorise public release of actual records. The worksheet is a disposition framework only. Real conditions require appropriate evidence and authority. The positive monetary outputs remain conditional on an accepted admissible menu and consequence perimeter.

Decision implications

- A positive expected monetary result does not establish permission or adequate safety evidence.
- Mandatory and public research purposes require their own explicit basis.
- Unresolved consequential boundaries should remain visible in acceptance.

GOVERNANCE AND APPLICATION

20

The strongest argument for broad exploratory research

An explicit boundary of the thesis

Exploration can change the decision that the initial model assumed

The thesis that information creates value through a better decision needs an important scope condition. The fixed-menu model assumes the institution knows its possible actions and represented states before commissioning the signal. Exploratory research may discover an action, a state, a mechanism or a capability that was absent from that architecture. Its contribution can be substantial, but the existing EVSI does not measure it. A paper that treats the fixed binary calculation as a complete research programme valuation would overextend its own method.

Exploration can begin with an explicit discovery purpose. A team may seek feasible operating designs, identify unknown interfaces or map a problem whose consequences are not yet well defined. The commission should describe the evidence object it expects and the decision that will follow its discovery stage. It need not invent a precise probability of finding a useful idea merely to fit the current app. A bounded budget, milestones and challenge process can make the programme accountable while acknowledging that the action set is endogenous to the research.

The distinction between learning about a state and creating an option should survive publication. A study that discovers a viable smaller design changes the menu; a test that estimates whether an existing design will work changes a belief within a menu. Both can improve decisions, but their numerical valuation requires different structures. The paper's

restricted-action example removes an option explicitly and demonstrates the menu effect. It does not calculate how likely research is to create that option or how much an option-discovery programme should cost.

An exploratory programme can use staged decisions without claiming that a single Stage row models them. A first phase may produce an evidence map, a subsequent phase may investigate a bounded mechanism, and a later decision may compare actual options. Each phase should have its own purpose, owner and acceptance condition. The institution should review whether the next stage is still warranted once discoveries change the problem. This is a governance framework for staged exploration. A quantitative dynamic programme would need an implemented model of discovery outcomes, costs and later choices.

The institution should avoid the opposite overreach of declaring exploration beyond accountability. An undefined promise that research may reveal something valuable can justify indefinite work if there is no reviewable object or decision path. The charter can state what counts as progress, which knowledge gaps matter, what evidence would narrow or end the programme and who can make that judgement. These conditions can be qualitative when numerical support is absent. Their purpose is to make learning responsive to a legitimate need rather than force unsupported precision.

Challenge a hypothetically narrow menu before buying another state signal. A supplier-led commission could compare only its proposed technology while operating redesign avoids the need; an investment screen could omit a feasible middle option between commitment and holding. These are design risks, not observed organisational cases. An independent challenge should seek an admissible alternative that could falsify the recommendation, because repairing the menu may improve the decision before further evidence is purchased.

Exploration may reveal that the binary state itself is unsuitable. Partial success, a new failure mechanism or an interacting condition can require a richer representation. The team should then revise the state and consequence account explicitly. It should not force the discovery into a success or failure label simply to preserve the old calculation. The fixed fixture remains useful as a record of the initial screen, while the new evidence defines the next model. A corrected architecture is a meaningful research result even if it prevents a neat continuity of metrics.

Capability building is another distinct purpose. Staff may learn a method, an institution may establish measurement access, or an operating team may develop the ability to implement a future response. Those benefits

can influence later decisions, but they are not identical to current state information. The commission should identify the capability, its use and acceptance conditions. The paper does not estimate a general monetary return on training or research capacity. A programme claiming such a value needs its own consequence model and evidence rather than an unbounded narrative benefit attached to every study.

Public knowledge may also have value beyond the commissioning institution. A shared dataset or method can inform many users whose decisions are not in the original payoff table. This can justify a public research purpose under an appropriate mandate and appraisal. It should be described with its audience, accessibility, quality and expected use. A sum of hypothetical private information ceilings is not a substitute for that appraisal. Overlapping uses, distribution and uncertainty about adoption need to be considered if a quantitative programme valuation is attempted.

Exploratory results should be reported as discoveries with their support and limits. A possible option may still need technical, commercial or institutional feasibility evidence before entering an operative optimiser. A newly identified state may lack a probability estimate. A promising mechanism may be an hypothesis rather than a measured effect. The institution should preserve those distinctions. Adding every idea as an available high-return action can overstate value just as surely as omitting every new alternative can understate it.

The next decision should be explicit. A discovery stage might justify accepting an option for detailed design, commissioning a causal test, declining a path or redefining the objective. The institution can then construct an appropriate model for that decision with better specified actions and evidence. The current binary app is useful when the revised problem genuinely has a fixed two-state and single-signal structure. It is not a universal endpoint toward which every exploratory programme must be compressed. Method selection should follow the problem that actually emerges.

The programme should retain a falsifier. What finding would make the proposed discovery path unhelpful, unfeasible or dominated by another approach? What would show that the expected capability is not being created? What would close the relevant knowledge gap sufficiently to act? These questions prevent a programme from redefining every outcome as progress. A qualitative falsifier can be rigorous when it refers to a concrete evidence object and decision. It should not be an invented statistical claim where the exploration has no defined sampling or probability model.

An exploratory finding should be classified by what it changes. A hypothetical discovery of a compatible alternative adds an action and requires a new consequence account. Discovery of an unrepresented failure mechanism changes the state model. A usable measurement method can improve a later instrument but does not yet provide a state observation. A clarified institutional objective changes valuation rather than likelihood. These routes all create possible value, yet none is ordinary updating of the original binary prior. The discovery register should identify the changed object and the next method needed. This preserves exploration's contribution without attributing a numerical programme value to the single-signal engine.

Exploration also needs a transition from discovery to a discriminating decision. In a hypothetical option-search programme, a promising alternative can advance to a compatibility test; an incompatible option can be retired with its reason; an unresolved option can remain in a bounded enquiry if a feasible observation would change its status. Continuing broad search is a separate decision requiring its intended use and remaining burden. The charter should record these dispositions rather than treat every new report as evidence that exploration must continue. This offers accountability without freezing the action menu prematurely, and gives a later reproducible model a clear account of how its alternatives and state definitions emerged.

Evidence and scope

NASA decision analysis connects alternatives, assumptions and evaluated results with a recorded rationale.

Evidence: NASA Systems Engineering Handbook

Completed illustrative exploration-to-decision charter

No discovery rate, actual programme or realised option is invented.

Completed illustrative exploration-to-decision charter			
Exploratory object	Acceptance question	Possible next decision	Method boundary
Evidence map	Are consequential gaps and sources clear?	Select a bounded investigation	No empirical likelihood inferred
Candidate option	Is a response sufficiently defined to assess?	Commission feasibility or reject	Not yet an operative action
New failure state	Does it change the consequence perimeter?	Revise state model	Binary aggregation may fail
Proposed mechanism	What evidence could distinguish it?	Design an appropriate test	No causal effect yet established
Capability	Can the institution use it for a stated purpose?	Accept, revise or stop development	No general return rate calculated
Public evidence object	Can intended users inspect and apply it?	Publish through an authorised process	Broader value needs its own appraisal

Commission discovery with a next decision and a falsifier

State what the exploratory phase should produce, who will assess it and what decision follows. Preserve unresolved feasibility and probability support when the menu or state changes. Use bounded milestones and a concrete falsifier to maintain accountability. Construct the next quantitative model only when its action and evidence architecture is sufficiently defined.

1. Name the discovery object and intended use.
2. Distinguish new options from state observations.
3. Specify acceptance and stop conditions.
4. Select the next method from the revised problem.

Discovery is difficult to specify before it occurs

An exploratory commission cannot list every insight in advance. It can still define the problem, relevant knowledge gaps, use conditions and review authority. Milestones can assess whether the work is narrowing a consequential uncertainty or producing a feasible option, while allowing unexpected findings to be reported honestly. Accountability need not require a false claim that discovery is fully predictable.

Endogenous actions and research programmes are outside the engine

The app assumes fixed actions, binary states and a single signal. It does not calculate option discovery, capability returns, public research benefits or a dynamic exploration policy. The charter is an original conceptual framework. Quantitative claims about those purposes require a separately specified and reproduced method with supported inputs.

Decision implications

- Discovery can change the action set or state model rather than only a belief.
- Exploration still needs an intended use, review authority and falsifier.
- The next method should fit the decision that actually emerges.

PRACTICAL APPENDICES

21 Appendix: probability and value dictionary

A replication procedure

The same label must mean the same thing in every object

The reproducibility dictionary connects a decision term to a specific computational field and meaning. Its purpose is to prevent a reader from treating similar labels as interchangeable. Prior probability, instrument sensitivity, specificity, branch mass and posterior success probability are all probabilistic quantities, but they condition on different information. Expected payoff, weighted branch contribution, gross information value and net information value are monetary quantities with different perimeters. A complete reproduction object preserves these distinctions rather than relying on a summary metric alone.

The input fixture is the supplied decision problem. It contains a prior, conditional signal likelihoods, an aggregate study cost and the action payoff cells. The fixtures here are explicitly illustrative. None is a fitted empirical object. The engine validates the input structure, applies its declared binary risk-neutral method and returns a structured output. The claim ledger then identifies each scalar selected for publication. A reader should interpret a displayed number through that chain, not as a free-standing fact whose meaning survives every change of context.

The prior field is displayed in percentage units on input, while branch mass and posterior values are stored as fractions on output. The binding format can display those fractions as percentages. This is a presentation transformation, not a revised belief. The full JSON remains the authoritative computational object. A reader reproducing a figure should

use its selected raw output pointers and the stated plot scale. Multiplying a percentage input again as though it were a fraction would change the model, so units belong in both the dictionary and the exhibit.

The action cells are aggregate monetary payoffs in USDk. The app does not infer a duration, exchange rate, discount schedule or cost incidence from that label. Those meanings must be supplied through the payoff perimeter. The same applies to testCost: the engine subtracts the aggregate amount once from gross information value but does not assemble its vendor, internal, waiting or operating components. An actual dossier should retain the separate evidence behind that aggregate. The fixture's transparent simplicity is not an empirical certification of every input component.

Input dictionary

Input names refer to the actual canonical schema; empirical provenance remains an independent obligation.

Input dictionary			
Field	Meaning	Unit or type	Critical scope
prior	Probability of represented success before signal	Percent	Specified event and information set
sensitivity	Positive signal conditional on success	Percent	Supplied intended-use likelihood
specificity	Negative signal conditional on failure	Percent	Supplied intended-use likelihood
testCost	Aggregate acquisition cost	USDk	Separate from pre-test action payoffs
actions[].name	Available response label	Text	Admissibility assumed for fixture
actions[].success	Action payoff if success	USDk	Aggregate declared perimeter
actions[].failure	Action payoff if failure	USDk	Aggregate declared perimeter

The actions output is ranked by current expected payoff, not preserved in the original input order. The regret output preserves a separate action account. A scalar pointer therefore needs to identify the actual structured location and its meaning; an author should not assume that a row index always names the same action in every output. The paper's binding files select values from the named fixture and the tables show action readings alongside them. If an input changes the ranking, the output and narrative must be checked again.

The branches output separates the signal name, probability mass, posterior, chosen action, conditional expected payoff and weighted contribution. The conditional payoff is evaluated after the signal. Its weighted contribution enters the ex ante expected payoff with information. These are not additive without the branch weights. A rare branch can have an attractive conditional value and a smaller contribution to the full commission. The dictionary makes the weighting step visible so that a favourable branch is not presented as though it were the certain payoff of buying a study.

The perfect-information account uses the state-specific best payoffs from the same menu and the same prior. It is not a stronger estimate of the actual study's performance. The resulting ceiling can screen a proposed purchase under the declared scope, while the signal-specific gross gain requires the branch policy. Both subtract the same current best expected payoff. A reader should inspect that benchmark when comparing the figures. If different values or action menus have been used, the apparent relationship between gain and ceiling may no longer describe one coherent decision.

Net information value subtracts the supplied study cost from gross signal value. A positive result is a conditional purchase finding under the stated risk-neutral objective, not an authorisation of the eventual operating action. A negative result can coexist with a useful instrument or a legitimate broader research purpose. The dictionary should retain those distinctions in the field descriptions. The summary metric becomes more useful when the reader knows exactly which part of the institution's commissioning question it answers.

The regret fields compare each action with the state-specific best available payoff. Expected regret uses the same prior weighting. The app does not convert regret into a psychological scale, determine a maximum-regret policy or elicit utility preferences. Publishing those labels without their scope could suggest features that the engine lacks. The table here identifies them as outputs of the fixed monetary matrix, and the conceptual preference chapter describes the additional methods that might be needed for a real institutional rule.

The priorSweep output is a sampled current-policy account. It varies the prior through the engine's predefined points and reports the best action and value at each point. It does not vary signal performance, calculate EVSI at every point or assign probabilities to alternative priors. The thresholds output contains pairwise equality points plus the complete-menu best action at each point. These are candidate comparisons under

linear state payoffs. The dictionary prevents the plot from being mislabelled as a continuous signal-value curve or every pairwise point from being called a policy transition.

The complete output contains display-oriented metrics, tables, warnings and method descriptions alongside decision-relevant values and the exact rational evaluation. These fields support the public app interface. The paper's quantitative claims bind to the named fixture and corresponding domain values. Duplicate representations of a number in a display table are not independent evidence. If a narrative relies on a qualitative warning, check whether its stated conditions hold in that fixture. In the no-policy-change case, exact EVSI is zero because the implemented policy remains unchanged. The stored exact fractions establish that arithmetic identity; converted numeric displays support reading without creating a different computational result.

Selected output dictionary

The field meanings follow the actual implementation; the limitations travel with them.

Selected output dictionary			
Field	Meaning	Unit	Does not establish
evWithout	Best current expected payoff	USDk	Realised outcome or permission
evWithInformation	Weighted conditional-policy payoff	USDk	Guaranteed study benefit
evPerfect	State-specific best policy expectation	USDk	Attainable perfect instrument
evpi	Perfect-information gain	USDk	Approval of every cheaper study
evsi	Gross gain from supplied signal	USDk	Empirical instrument validation
netInformationValue	Gross gain less supplied cost	USDk	Complete cost assembly
branches[].probability	Signal branch mass	Fraction	Frequency estimated from actual data
branches[].posterior	Conditional success probability	Fraction or unavailable	A confidence score
branches[].weightedValue	Conditional payoff weighted by branch	USDk	Standalone branch outcome
thresholds[].prior	Pairwise equality prior	Percent	Always an active policy switch

The bindings file supplies the bridge from a published claim token to its input or output pointer. It retains the basis, model owner and associated ledger entry. A displayed claim should have reciprocal references to the block in which it is used. A selected plot coordinate needs the same link even when its value is not printed in the caption. This makes the figure reproducible from the declared object. It also prevents a chart from silently using a calculation different from the table beside it.

The dictionary's final category is unimplemented method. Causal estimation, utility elicitation, joint evidence, dated delay, adaptive stopping, implementation performance and research programme valuation are discussed where relevant. They are not domain outputs of the binary app. A real dossier needing one of them should specify its method and evidence before making a numerical claim. The presence of a chapter or worksheet does not substitute for a computation. This boundary is part of the reproduction object because it states what the object cannot honestly demonstrate.

Use the dictionary to trace a suspected interpretation error to its conditioning event. A hypothetical reviewer comparing the positive branch posterior with the unconditional prior is comparing different quantities; one is conditioned on the signal and the other is not. Comparing branch mass with sensitivity also mixes a marginal signal probability with a state-conditional likelihood. Comparing weighted branch contribution with conditional expected payoff omits the branch's occurrence mass. These errors can survive immaculate units and accurate copying. The field map should identify the conditioning statement before a value is transferred to prose, a chart or a response rule, then retain the complete input and output identity for the comparison.

Evidence and scope

Sheffield describes TSD guidance on selecting evidence for model structure and parameters; only its publisher description is used.

Evidence: Reviewing model parameters: NICE Decision Support Unit technical support documents

Use the field, fixture and scope together

Identify the exact computational field before copying a value. Retain its units, conditioning and fixture identity, then verify the reciprocal claim binding. Treat display-oriented duplication as another representation of the same result. Require an additional specified method for any quantity outside the dictionary rather than inferring it from an action name or chapter title.

1. Read the field meaning and units.
2. Select the named fixture and exact pointer.
3. Preserve raw values and declared display format.
4. State the missing method when scope expands.

A reader should not need to understand raw pointers

The main argument and labelled tables should be intelligible without inspecting software objects. The pointer desk serves the motivated reviewer who wants exact reproduction. Layered access lets a practical reader use the decision logic while preserving a precise route for assurance. Simplicity of presentation should not remove the underlying object needed to verify it.

A dictionary records meaning rather than empirical validity

The field descriptions establish the intended implementation and publication scope. They do not validate a real probability, action payoff or study instrument. A reader applying the method to an actual opportunity must supply the corresponding evidence, authority and limitations. The complete frozen illustration remains a methodological object.

Decision implications

- Conditioning and units distinguish superficially similar fields.
- Ranked outputs require exact fixture and pointer identity.
- A method discussed conceptually is different from an implemented output.

PRACTICAL APPENDICES

22 Appendix: illustrative information studies

Full input/output ledgers and sensitivity

Distinct mechanisms with complete, traceable inputs

The worked studies isolate distinct mechanisms that can change an information purchase. Their inputs are authored illustrations and their outputs come from the same canonical Decision lab engine as the public app. They are not sampled projects, observed rates or estimates fitted to client records. The studies retain complete fixture, output and binding

files in the replication desk. The tables below publish the decision-relevant input and result account so that a reader can understand the comparison before inspecting the full structured object.

The original action matrix is shared deliberately across several studies. This lets a change in the prior, signal likelihoods or acquisition price be examined while holding the response consequences fixed. The restricted-action study removes Commit; the reduced-upside study supplies a different matrix. Those are distinct decision-perimeter changes. The appendix shows the shared matrix once, identifies its applicable cases and preserves the alternative matrices separately. This avoids publishing repeated identical tables while keeping every fixture fully reproducible.

The useful-signal study provides the baseline mechanism. Stage is the best current action. A positive result selects Commit and a negative result selects Hold. The instrument therefore changes the conditional policy and creates a gross expected monetary gain. Its supplied acquisition cost leaves a positive net result. This finding is conditional on the prior, likelihoods, payoff perimeter and feasibility assumptions. It is not an empirical statement that a particular commercial study has that value or that a committee should make an actual commitment.

The no-policy-change study raises the prior while retaining the original matrix and instrument. Commit is already best and remains best in both branches. The signal updates belief but does not improve the implemented fixed-menu policy. Its gross gain is exactly 0.00 USDk in the current rational evaluation; the complete file preserves that equality and the numeric display. The supplied positive price therefore produces a negative net result. This case falsifies an unqualified claim that an informative or accurate signal always justifies a new research purchase. Its scope remains the supplied menu, state model and implementation conditions.

The expensive-signal study changes the price rather than the information mechanism. Its conditional actions and gross gain match the useful baseline because the prior, likelihoods and matrix are unchanged. The larger supplied price makes the net result negative. This case distinguishes the methodological usefulness of an instrument from the economic case for buying it. A committee could negotiate, redesign the scope or decline the commission, but the fixture does not model those procurement responses or infer the supplier's actual cost.

The false-positive-burden study combines a low prior with a weak negative-state filter. The positive signal raises the success probability but not enough to make a paid exposure preferable under the original matrix.

Both branches select Hold. This route to redundancy differs from the already-favourable no-change case. It illustrates why a signal's likelihoods and the prior must be read together against the action boundary. A favourable direction of update does not by itself establish a sufficient condition for commitment.

The inverse-signal study deliberately contradicts the usual intuitive meaning of positive. Under its supplied likelihoods, a positive result lowers the success probability and selects Hold, while a negative result selects Commit. The signal still has positive net information value in the fixture because the conditional policy uses its actual association with the state. The case tests whether the reader follows likelihoods rather than labels. It does not validate a real negatively associated instrument or recommend using one without a relevant performance record.

The perfect-price-boundary study makes the signal perfectly informative within the binary illustration and supplies a price equal to its gross gain. Its net result is at the model boundary. The positive branch selects the success-state best action and the negative branch selects the failure-state best action. The case demonstrates the ceiling relationship exactly within the supplied arithmetic. It should not be read as evidence that perfect commercial observation is available at that price, or that a purchase with no margin for unresolved cost is prudent.

The restricted-action study removes Commit from the menu and retains the original likelihoods, prior and remaining payoff cells. The same instrument can now support Stage after a positive result and Hold after a negative result. The reduction in response opportunity lowers gross information value enough that the supplied price exceeds it. This case falsifies a purchase rationale that assumes an unavailable action. It does not calculate how to obtain authority, the chance of receiving approval or the value of an actual governance change.

The reduced-upside study keeps the instrument and prior but supplies smaller success payoffs for the active actions. The new matrix produces a different current value and information gain. It illustrates how an opportunity revision can require a new commission appraisal. No dated delay, discount rate or causal relationship is inferred. A real case linking study time to changed payoffs would need its own schedule and consequence evidence. The original and revised matrices are preserved so the reader can see exactly which assumption changed.

The branch account publishes mass, posterior and selected action for every study. Conditional expected payoff and weighted contribution are available in the full outputs; selected tables in the main argument show

that weighting step. The appendix retains the complete branch interpretation without copying every display-oriented table from the app. The replication desk remains the full reproduction object. A reader can verify any scalar against its binding and rerun the named fixture through the paired canonical engine.

The policy sweep depends on the action matrix rather than the fixture's initial prior or signal performance. The original-matrix studies therefore share the same sweep. The restricted and reduced-upside matrices have different policy accounts. The graphical comparison publishes those distinct sweeps, with points selected from their actual canonical outputs. It is a picture of best current expected payoff across sampled priors. It does not plot EVSI, assign plausibility to the priors or imply that an empirical probability distribution has been estimated.

Each study should be challenged through its decisive assumption. For the useful baseline, ask whether the branch responses are feasible and the likelihoods apply. For no change, ask whether another consequential action has been omitted. For high price, ask whether the cost boundary is complete. For false positives, ask whether the prior and filter are supported. For the inverse case, ask whether the signal meaning is understood. For the boundary, ask about unresolved cost. For restriction and reduced upside, ask whether the decision perimeter has actually changed in the manner supplied.

The study slate can therefore select the next enquiry. A no-change result asks whether an active action boundary is reachable; an expensive result asks whether the purchase burden can be reduced or its purpose changed; a restricted-menu result asks whether the response belongs in the feasible set. An inverse signal asks whether names match likelihood meaning, while reduced upside asks whether the consequence account still represents the opportunity. These routes cannot be cured by the same additional report. The cases remain fictional controlled mechanisms, and their value is to locate the missing evidence or intervention without borrowing their numerical assumptions as an empirical baseline.

Illustrative calculation. Assumptions and sensitivity are published with the model.

Original study cohort: illustrative instrument input account

Likelihood and prior fields are percentages. The aggregate acquisition price is USDk. Additional numerical boundary studies are reported separately.

Original study cohort: illustrative instrument input account					
Study	Prior	Sensitivity	Specificity	Cost, USDk	Action matrix
Useful signal	40.00	85.00	85.00	5.00	Original
No policy change	95.00	85.00	85.00	5.00	Original
Expensive signal	40.00	85.00	85.00	80.00	Original
False-positive burden	8.00	90.00	50.00	4.00	Original
Inverse signal	40.00	20.00	30.00	5.00	Original
Perfect price boundary	5.00	100.00	100.00	9.00	Original
Restricted action	40.00	85.00	85.00	5.00	Restricted
Reduced upside	40.00	85.00	85.00	5.00	Reduced upside

Illustrative calculation. Assumptions and sensitivity are published with the model.

Complete distinct illustrative action matrices

The original matrix applies to the original-matrix studies named in the instrument table. All values are USDk.

Complete distinct illustrative action matrices			
Matrix	Action	Success payoff	Failure payoff
Original	Commit	180.00	-100.00
Original	Stage	70.00	-15.00
Original	Hold	0.00	0.00
Restricted	Stage	70.00	-15.00
Restricted	Hold	0.00	0.00
Reduced upside	Commit	100.00	-100.00
Reduced upside	Stage	30.00	-15.00
Reduced upside	Hold	0.00	0.00

Illustrative calculation. Assumptions and sensitivity are published with the model.

Original study cohort: illustrative current and conditional-policy result account

All monetary values are USDk. The complete files retain exact rational evaluations alongside numeric display values. Reporting precision does not redefine algebraic equality or purchase sign. Additional numerical boundary studies are reported separately.

Original study cohort: illustrative current and conditional-policy result account					
Study	Current action	Current value	Signal policy value	Gross gain	Net gain
Useful signal	Stage	19.00	52.20	33.20	28.20
No policy change	Commit	166.00	166.00	0.00	-5.00
Expensive signal	Stage	19.00	52.20	33.20	-46.80
False-positive burden	Hold	0.00	0.00	0.00	-4.00
Inverse signal	Stage	19.00	39.60	20.60	15.60
Perfect price boundary	Hold	0.00	9.00	9.00	0.00
Restricted action	Stage	19.00	22.45	3.45	-1.55
Reduced upside	Stage	3.00	25.00	22.00	17.00

Illustrative calculation. Assumptions and sensitivity are published with the model.

Original study cohort: illustrative perfect-information account

The same fixture perimeter supplies the state-specific best policy and its gain over the current action. Additional numerical boundary studies are reported separately.

Original study cohort: illustrative perfect-information account		
Study	Perfect policy expected payoff, USDk	Perfect-information ceiling, USDk
Useful signal	72.00	53.00
No policy change	171.00	5.00
Expensive signal	72.00	53.00
False-positive burden	14.40	14.40
Inverse signal	72.00	53.00
Perfect price boundary	9.00	9.00
Restricted action	28.00	9.00
Reduced upside	40.00	37.00

Illustrative calculation. Assumptions and sensitivity are published with the model.

Original study cohort: illustrative signal branch account

Mass and posterior are displayed as percentages. Weighted contributions are USDk. Additional numerical boundary studies are reported separately.

Original study cohort: illustrative signal branch account					
Study	Signal	Branch mass	Posterior success	Selected action	Weighted contribution
Useful signal	Positive	43.00%	79.07%	Commit	52.20
Useful signal	Negative	57.00%	10.53%	Hold	0.00
No policy change	Positive	81.50%	99.08%	Commit	144.60
No policy change	Negative	18.50%	77.03%	Commit	21.40
Expensive signal	Positive	43.00%	79.07%	Commit	52.20
Expensive signal	Negative	57.00%	10.53%	Hold	0.00
False-positive burden	Positive	53.20%	13.53%	Hold	0.00
False-positive burden	Negative	46.80%	1.71%	Hold	0.00
Inverse signal	Positive	50.00%	16.00%	Hold	0.00
Inverse signal	Negative	50.00%	64.00%	Commit	39.60
Perfect price boundary	Positive	5.00%	100.00%	Commit	9.00
Perfect price boundary	Negative	95.00%	0.00%	Hold	0.00
Restricted action	Positive	43.00%	79.07%	Stage	22.45
Restricted action	Negative	57.00%	10.53%	Hold	0.00
Reduced upside	Positive	43.00%	79.07%	Commit	25.00
Reduced upside	Negative	57.00%	10.53%	Hold	0.00

Replicate the mechanism before transferring the conclusion

Open the named study input and paired output in the replication desk, rerun the paired app engine and inspect the selected bindings. Identify the decisive assumption that differentiates the case. Transfer the method only after supplying evidence for the real event, likelihoods, payoffs and admissibility. Preserve a changed opportunity as a new decision record rather than silently modifying a published illustration.

1. Select a named frozen fixture.
2. Compare full raw output and displayed bindings.
3. Challenge the case-specific mechanism.
4. Replace assumptions only with a documented real-decision perimeter.

A compact case set cannot represent all institutions

The studies are designed counterexamples and mechanism demonstrations, not an empirical sample or a completeness claim. Additional states, instruments and decision rules can change the conclusions. Their value is to make the fixed binary logic inspectable and reveal specific failure conditions. A reader should choose or build a method that fits the actual decision rather than infer universality from the number of illustrative cases.

The complete reproduction object is larger than the selected tables

The full output files retain all domain and display values, warnings, sampled policies and pairwise intersections. The appendix publishes the decision-relevant account and avoids repeated identical sweeps. No actual data, client cases, study performance observations or legal permissions are inferred. The model limitations and fixture assumptions govern every numerical reading.

Numerical equality and a supported change of scale

Illustrative calculation. Assumptions and sensitivity are published with the model.

Economic break-even should describe an equality under the stated assumptions. The three price-precision fixtures retain the default payoff matrix, prior and likelihoods and change only acquisition price. The exact-price case sets cost 94.99 USDk and yields net information value 0. The nearby prices are 94.989999999 and 94.990000001 USDk. Their computed net values are 1e-09 and -1e-09. The first is positive and the second negative; neither is an exact break-even.

Joint signal-state masses and action payoffs are now evaluated as rational quantities derived from canonical decimal inputs. Branch actions are ranked before any division or numeric display, so posterior rounding cannot change a preferred response. The exported evaluation records exact EVSI, EVPI and net sign. This guarantee concerns arithmetic alone. A surplus of a billionth of USDk is not credible economic precision when probabilities, payoffs, permission or delay remain uncertain. A decision

owner should challenge those assumptions and agree a substantive decision margin rather than let a software tolerance silently redefine equality.

Decision implications

- The studies are designed mechanism tests with fully retained reproduction objects.
- Shared matrices are published once while distinct perimeters remain visible.
- Transfer requires evidence and authority beyond numerical reproduction.

PRACTICAL APPENDICES

23 Appendix: research commission worksheets

Practical decision-relevant research templates

From a conditional calculation to an accountable commission

The evidence supplier, analyst and action owner should recognise the same decision in the commission. These completed illustrative dispositions use the useful-signal mechanism as a bounded reference. Empirical likelihoods, actual payoffs, legal authority and implementation readiness remain unresolved. Adaptation therefore requires evidence and accountable approvals, rather than copying the hypothetical purchase conclusion. A filled record makes that work inspectable; a blank form or hypothetical result does not complete it.

The first task is to distinguish a decision question from a research topic. A topic such as whether a proposed activity is promising can support almost any study. A decision question identifies the currently available actions, the represented state, the response window and the consequence of choosing incorrectly. The completed charter asks whether to purchase a binary signal for a specified Commit, Stage or Hold menu. That wording makes it possible to challenge a study that delivers an interesting description without supporting the conditional response policy the institution needs.

The charter should also preserve the current counterfactual. If the institution already intends to Commit irrespective of the signal, buying a study cannot be justified by comparing its favourable outcome with Hold.

If the current preferred action is Stage, the signal policy must improve on Stage under the declared objective. The fixture provides that explicit benchmark, but an actual institution must establish what would happen without the proposed evidence. A sponsor who disagrees with the current action should challenge the matrix or admissibility before using an inconsistent counterfactual in the study valuation.

The event register fixes what the represented success state means. A commercial outcome, a technical threshold and a regulatory permission can be different events even when a proposal calls all of them success. The illustrative binary state in the app is deliberately abstract; the commissioning worksheet requires a real decision to define its event, horizon, measurement and information set. If a finding is used to update a different event, the numerical posterior no longer follows merely because the same word appears in both reports. The event owner should confirm that bridge explicitly.

The instrument register asks for conditional likelihood support. It is insufficient for a supplier to state an overall accuracy percentage without the intended event, reference standard and use conditions. The app needs a positive-signal likelihood given success and a negative-signal likelihood given failure. For a real commission, those inputs require evidence, uncertainty and transfer limits. The worksheet therefore separates the supplied illustrative likelihoods from the unresolved empirical validation. It can support a scoped mechanism demonstration while refusing to certify that an actual instrument has the fixture's performance.

The payoff record should show why each action cell belongs in one matrix. An action can have an upside in success and a burden in failure, but those amounts need a shared baseline, horizon and institutional incidence. A reversible Stage action may retain benefits and avoid losses differently from a full commitment. The fixture represents these consequences in aggregate; it does not explain a technology, schedule or contract behind the Stage label. The worksheet asks the action owner to supply that explanation and to confirm that the choice can actually be taken in both states.

Study cost is a separate commissioning perimeter. A quoted vendor fee may exclude internal review, operational disruption, interpretation and waiting burdens. Conversely, an operating consequence already inside an action payoff should not be counted again merely because the institution buys evidence. The completed cost record requires a component owner, treatment and timing statement. It declines to populate real amounts.

The aggregate fixture cost is sufficient for the illustrative subtraction, while an actual purchase needs its own complete acquisition account and a check against duplication.

The response protocol is the practical link between a posterior and a decision. It should identify the signal format, the action proposed for each branch, the authority needed to implement it and the treatment of an unusable delivery. A result that arrives after an irreversible commitment may have no value for that commitment even if it remains technically informative. A result that requires authority the institution lacks can encounter a similar implementation gap. The worksheet records these as feasibility obligations. It does not hide them in a probability adjustment with no defined empirical meaning.

The acceptance specification should state what the supplier must deliver for the decision use. This might include a defined event, conditional performance evidence, uncertainty description, full data provenance, a method that can be inspected and a timely interpretable output. The actual scope depends on the commission. The completed example requires a binary classification object appropriate to the represented event and a transparent bridge into the model inputs. It does not demand a universal report format. Acceptance should be based on whether the contracted evidence object supports the declared use, not whether the supplier repeats the committee's preferred conclusion.

A challenge worksheet makes the recommendation conditional in a useful way. It identifies a premise whose admissible change could alter the commissioning decision, the evidence needed to resolve it and the interim action. For this example, the response menu, likelihood transfer, cost perimeter and usable window are decisive candidates. The paper's designed counterexamples show different ways these premises matter. An actual challenge should avoid claiming that a chosen stress value is probable unless there is support for that probability. The purpose is to expose dependence on an assumption, then commission or collect the specific evidence that can resolve it.

The minute should record each judgement under its owner: the analyst reproduces the frozen canonical output; the method reviewer checks the stated binary equations; the action owner confirms feasibility and mandate; the evidence owner defends empirical inputs. None supplies the others. Retain unresolved conditions beside the illustrative result so a computational signature cannot later be read as certification of an actual research purchase.

The interim disposition is sometimes the most useful part of the record. If the instrument's performance is unsupported, the institution can ask for a validation proposal, choose a narrower pilot or decline the purchase. If the response is infeasible, it can repair the action menu or change the research purpose. If cost is incomplete, it can request a full acquisition account before comparing a price with information value. These responses turn uncertainty into a concrete work programme. They are more useful than treating every incomplete input as an invitation to select a convenient point estimate and proceed.

The worksheets also help structure a supplier discussion. The client can explain the decision event, intended responses and evidence acceptance conditions without asking the supplier to promise a favourable policy outcome. The supplier can challenge an impossible likelihood request, an unsuitable reference standard or a response window that cannot be met. The analyst can then revise the model perimeter before a commercial commitment. This is a substantive negotiation about what evidence is usable. The binary fixture can clarify incentives and thresholds, but it cannot establish the actual contract terms or resolve a dispute about applicable duties.

After delivery, a closeout record should compare the accepted evidence object with the commissioning purpose. It should preserve the finding, input changes, selected action, authority and material departure from the protocol. A disappointing signal is not automatically supplier failure; an unusable or unsupported delivery is a different question. Likewise, an unfortunate outcome does not by itself prove that a coherent expected-value decision was mistaken. The institution should separate method error, input error, implementation failure and uncertainty that remained within the accepted scope. This classification allows learning without rewriting the original record around hindsight.

Repeated commissions could support a separately designed study of delivery quality, response implementation and calibration. A successful-case archive alone cannot establish representative frequencies. Declare the collection universe, inclusion rules and comparison method before an empirical claim. This paper's illustrations have no such universe; future evaluation may use the original decision records as traceable inputs, but cannot treat them as proof of organisational improvement.

Follow a hypothetical capacity gap through the filled record. An acceptable instrument specification leaves purchase held when the favourable-result response lacks a usable resource reservation. The response owner records missing capacity; the interim minute retains feasible Stage or Hold; the handoff requires capacity work rather than

extra likelihood precision. If capacity becomes available, rerun the current menu and acquisition account before progression. The completed charter exposes the gap without supplying authority or field evidence.

Completed illustrative research commission charter

The illustrative useful-signal mechanism supplies a reference calculation. No actual purchase or institution is represented.

Completed illustrative research commission charter		
Record item	Illustrative disposition	Required real-decision completion
Decision question	Purchase a signal for Commit, Stage or Hold under the declared monetary objective	Name the actual commitment and feasible alternatives
Represented event	Binary success or failure, defined only for the illustration	Define the event, horizon and measurement
Current counterfactual	Use the current best expected-payoff action from the named fixture	Confirm the action that would occur without evidence
Use of delivery	Condition the response on the accepted signal branch	Approve a timely branch protocol and fallback
Institutional owner	No actual owner claimed	Name payoff, funding and action authorities
Disposition	Illustrative purchase finding remains conditional	Resolve material evidence and feasibility gaps before action

Completed illustrative input provenance and transfer register

The named fixture supplies transparent assumptions. These entries do not turn them into empirical estimates.

Completed illustrative input provenance and transfer register		
Input	Current status	Evidence or review needed
Prior	Explicit scenario assumption	Event-specific evidence and information-set review
Conditional signal likelihoods	Explicit scenario assumptions	Intended-use validation and uncertainty account
Commit payoff cells	Aggregate scenario amounts	Commercial, operating and incidence evidence
Stage payoff cells	Aggregate option labelled Stage	A feasible operating design and complete consequences
Hold payoff cells	Declared reference amounts	Baseline and foregone-opportunity review
Study cost	Aggregate illustrative acquisition amount	Vendor, internal, timing and disruption cost account
Available menu	Illustrative admissibility	Mandate, constraint and implementation review

Completed illustrative acquisition-cost scope record

No real prices or cost observations are supplied. Each item has a declared treatment to prevent omission or double counting.

Completed illustrative acquisition-cost scope record		
Component question	Illustrative treatment	Owner task before purchase
Supplier delivery price	Included only through the supplied aggregate cost	Confirm scope, price basis and acceptance object
Internal review and interpretation	No independent amount estimated	Document material workload and valuation basis
Waiting consequences	No dated calculation implemented	Establish usable window and consequence perimeter
Operating disruption from testing	No actual disruption asserted	Assess whether action payoffs or cost account includes it
Repeated or follow-up evidence	Outside the static purchase illustration	Specify dependency and incremental programme method
Costs already incurred	Exclude from the remaining purchase comparison where unavoidable	Record sunk versus remaining avoidable commitments
Duplicate consequence	Do not add a burden already inside action payoffs	Reconcile the full matrix and acquisition account

Completed illustrative response and acceptance protocol

A branch policy is useful only if an accepted delivery can support a feasible response. This protocol contains no actual authority.

Completed illustrative response and acceptance protocol		
Condition	Illustrative response record	Real implementation requirement
Accepted positive signal	Use the named conditional-policy branch	Confirm action authority, capability and timely commitment
Accepted negative signal	Use the named conditional-policy branch	Confirm permission to decline or stage after an adverse finding
Unusable or missing delivery	Do not invent a posterior from an unaccepted signal	Apply an approved fallback and supplier-remedy process
Delivery outside usable window	Reassess the remaining decision	Identify commitments already irreversible
Evidence conflicts with prior inputs	Preserve conflict and revise through accountable review	Route event, likelihood and consequence disputes to owners
New action becomes feasible	Create a revised menu and reproduction object	Review new consequences and admissibility before comparison
Protected condition unresolved	Withhold unsupported monetary disposition	Obtain the appropriate authority or broader method

Completed illustrative challenge and interim-disposition register

Each challenge tests a consequential premise. None asserts a frequency for the stress condition.

Completed illustrative challenge and interim-disposition register		
Challenge	Evidence that could resolve it	Interim disposition
Signal does not change a feasible action	Inspect full branch policy and implementation authority	Do not justify purchase from precision alone
Intended-use likelihoods do not transfer	Obtain event-specific validation and transfer limits	Retain a scoped illustration, withhold empirical endorsement
Total acquisition burden exceeds gain	Complete the acquisition-cost perimeter	Compare only after costs and payoffs are reconciled
Available menu is restricted	Confirm which actions can be taken after delivery	Rerun the actual admissible menu
Opportunity changes before delivery	Specify remaining consequences and usable window	Use a revised decision record or appropriate timing method
Mandatory or public-purpose requirement applies	Obtain exact authority and purpose review	Assess the commission on its legitimate declared basis
Another signal shares the same evidence	Map common data and conditional dependence	Require a joint or incremental method before adding gains

Completed illustrative committee minute and closeout account

The minute separates reproduction, empirical validation and institutional authorisation.

Completed illustrative committee minute and closeout account		
Account	Illustrative record	Condition for an actual signed disposition
Computational identity	Named fixture, frozen output and canonical app version retained	Independent rerun and complete selected-claim comparison
Method scope	Binary, risk-neutral monetary comparison	Confirm that the represented decision fits the method
Evidence status	Public guidance plus explicit scenario assumptions	Defend actual event, likelihood and consequence inputs
Institutional acceptance	No real mandate or purchase approval claimed	Record responsible authority and unresolved challenge treatment
Delivery outcome	No field delivery represented	Preserve accepted object, departures and response
Learning account	Designed mechanism demonstration only	Declare collection universe and comparison method before empirical claims

Commission evidence through a completed, challengeable record

Start with the action charter and current counterfactual. Trace the event, conditional instrument performance, payoffs and acquisition costs to owners and evidence. Approve a timely response and an acceptance object, then record decisive challenges and interim dispositions. Preserve the signed perimeter and computational identity through delivery and closeout.

1. Define the decision before specifying the study.
2. Separate scenario assumptions from validated inputs.
3. Confirm an executable response protocol.
4. Record unresolved challenges beside the recommendation.
5. Evaluate delivery without confusing outcome uncertainty with method failure.

Extensive records can slow a small research purchase

Scale the record to the decision. A low-consequence reversible commission can use concise entries and a limited review. A consequential irreversible response needs a stronger evidence and authority account. The requirement is to make decisive premises visible, not to maximise paperwork. A short charter can be rigorous when its boundaries are clear and the remaining uncertainty is proportionate to the action.

Completed illustrations do not complete an actual commission

The entries provide concrete dispositions without inventing field evidence, prices, legal conclusions or institutional approvals. The reader must supply the real decision perimeter and resolve consequential gaps. No worksheet computes joint evidence, delay, utility, adaptive stopping or realised research programme performance.

Decision implications

- A usable commission connects the event, evidence object and executable response.
- A filled record should expose unresolved evidence rather than conceal it.
- Closeout distinguishes delivery failure, implementation failure and accepted uncertainty.

PRACTICAL APPENDICES

24

Appendix: evidence scope and an admissible-response packet

Establish the admissible branch policy and consequence ownership before commissioning the instrument.

The scope boundary

This methodological candidate combines the declared binary model, public primary guidance and original institutional analysis. Its scenario inputs and exact paired-app outputs expose reproducible mechanisms. They form no empirical survey or sampled case set, and estimate neither study value frequency, actual instrument accuracy nor institutional response behaviour. Readers can challenge those mechanisms without treating their illustrative values as observed commission evidence.

The principal method boundaries remain substantive. The app represents a binary state, a supplied binary signal, an admissible action menu and risk-neutral monetary payoffs. It has no engine for utility elicitation, causal estimation, correlated signals, dated delay, adaptive stopping, operating implementation or public research programme valuation. Chapters discuss those issues to identify when a broader method is needed. A conceptual framework or completed worksheet does not quietly extend the numerical calculation. An institution needing an additional quantity should specify and validate its method before publishing that quantity.

The source dates also matter. The retained guidance reflects the accessed public editions and their original purposes. A jurisdiction-scoped or domain-specific document does not automatically impose a rule on every institutional reader. The paper uses primary material for narrow methodological support and states those limits. A reader applying the method to a real decision should check the current applicable instrument, mandate and technical standard. This requirement concerns the actual decision scope; it does not turn the public methodological candidate into a legal or investment opinion.

Source cards support challenges to narrow institutional statements; canonical study files support challenges to illustrative inputs and outputs. The packet below then asks whether evidence changes an admissible

response, whose consequences count and who can implement it. Reproducible fictional arithmetic leaves actual-use authority and evidence unresolved.

A completed account and an executable disposition

The final commissioning packet should be a completed decision record rather than another restatement of the information-value contrast. The specimen below applies the already retained useful-signal and restricted-action studies to a fictional institutional situation. Its acquisition account is complete, its instrument specification is accepted and its current permission for Commit has expired. That combination makes the immediate decision intelligible: the sponsor has a priced instrument, but it lacks an executable favourable-result response. Every role, reference, allowance and state in this specimen is authored for illustration. None represents an ICM mandate, a real approval or an empirically validated likelihood.

Illustrative calculation. Assumptions and sensitivity are published with the model.

Completed fictional account: every commissioning cost has one place

Illustrative authored cost inputs in USD thousands. Fictional evidence references and accepted allowances are recorded in the retained commissioning specimen. The amounts are fictional inputs for the method illustration. The exact decimal component sum equals the existing canonical testCost.

Completed fictional account: every commissioning cost has one place			
Incremental component	Amount (USDk)	Account owner	Treatment
Vendor	3.0	Commission lead	Incremental instrument acquisition; charged once in testCost.
Interpretation	1.0	Instrument lead	Incremental analysis of the signal; charged once in testCost.
Disruption	0.5	Operations owner	Commissioning trial interruption only; charged once in testCost.
Waiting	0.5	Decision chair	Incremental wait before the signal; charged once in testCost.
Single aggregate test cost	5.00	Commission lead	The same aggregate is used once in both existing menu fixtures.

Illustrative calculation. Assumptions and sensitivity are published with the model.

Completed fictional disposition: an expired permission changes the commission

Illustrative authored records. The references name entries in the retained specimen, not real approvals or externally issued documents. The present disposition is complete even though a proposed permission renewal remains unresolved.

Completed fictional disposition: an expired permission changes the commission			
Record and owner	Current state	Expiry trigger	Present disposition
LAB-A / Instrument lead	Specification accepted	Population, threshold or instrument changes	Instrument can be commissioned under the stated likelihood assumptions; this record supplies no authority to Commit.
MENU-A / Decision chair	Permission expired	Delegation term ends or mandate changes	Commit is excluded now. A proposed renewal is pending and cannot support the current purchase appraisal.
CAPACITY-A / Operations owner	Stage and Hold confirmed	Resource allocation changes	The restricted action menu remains executable inside the decision window; scarce resources for Commit are unreserved.
PACKET-A / Commission lead	Purchase deferred	A material input or permission changes	Do not buy the signal at this account while Commit remains excluded. Reopen only after evidence and the model are reconciled.

The aggregate acquisition account is unchanged from the earlier studies. Vendor service, internal interpretation, commissioning interruption and waiting burden each occupy a separate component with a responsible role. The cost record assigns each item once to the proposed instrument purchase. It does not add the future cost of taking Commit, Stage or Hold. Those response consequences belong inside the corresponding action payoffs. This distinction lets a reviewer reconcile the complete account without charging an implementation burden once in testCost and again in every branch payoff. It also prevents an apparently complete vendor quotation from silently excluding internal effort or delay.

An accepted allowance is an evidence state within this fiction, not proof that the amount is accurate. The Commission lead accepts the instrument scope represented by QUOTE-A; the Instrument lead accepts the interpretation allowance in REVIEW-A; the Operations owner accepts the trial interruption in TRIAL-A; and the Decision chair accepts the commissioning wait in WINDOW-A. The specimen retains the component values and the exact decimal reconciliation to the canonical cost input. In a real commission, the same roles would need to obtain and challenge their own records. If waiting becomes prolonged or a trial changes its operating perimeter, the prior account expires and the aggregate cost must be recomputed before the purchase is reconsidered.

The evidence-state register distinguishes an accepted instrument from a current response permission. LAB-A accepts the fictional instrument specification. MENU-A records that the prior delegation has expired. The upcoming branch deadline has not passed. A renewal proposal is pending, so the packet excludes Commit from the current menu. CAPACITY-A confirms that Stage and Hold remain executable; it records no reserved resource for Commit. The Commission lead therefore defers the purchase under PACKET-A. This is a completed present disposition, rather than an incomplete claim that a positive unrestricted calculation can authorize action. The team may continue the permission work, but it cannot borrow the proposed future state to justify the instrument purchase today.

Reopening requires a change that the packet can actually evidence. If the Decision chair renews permission and the Operations owner confirms the required resource before the response deadline, the owner may restore Commit to the supplied menu and rerun the canonical appraisal with the current acquisition account. If permission remains expired, the restricted-menu reading continues to apply. A revised study price can also change that reading, but it must be recorded as a new cost account rather than retroactively changing the accepted allowance. An altered population, reporting threshold or instrument expires LAB-A as well; response authority cannot rescue a likelihood specification that is no longer applicable.

Hand over instrument scope, component ledger, admissible menu, branch recipients and expiry conditions together. A headline value conceals changed permissions or newly excluded burdens; a favourable result alone conceals an expired authority or resource window. The completed packet assigns each role a present responsibility and reopening

condition. It remains an illustrative binary, declared-monetary-utility, single-instrument method, supplying neither actual institutional authority nor validation of real records.

The specimen can route a failed reopening by its actual reference. If LAB-A remains current but MENU-A is not renewed, further instrument assurance does not restore Commit. If MENU-A is renewed while CAPACITY-A still provides no Commit resource, permission alone does not restore the response. If both are resolved but the response deadline passes, the team needs a current timing and consequence account rather than reuse the expired packet. These branches make the present defer disposition useful: each identifies the object that must change and the object that can remain accepted. They preserve the distinction between diagnostic, authority and operating responsibilities through the fictional record itself.

A restriction should not be given an invented probability merely to blend the complete and restricted calculations. MENU-A records a present expired delegation, not an uncertain future state of the opportunity. CAPACITY-A records a present response capability, not the instrument's sensitivity. A proposal to renew the delegation can be pursued separately with its own authority and burden, while the current purchase uses the presently executable menu. If a real restriction were instead legally unchangeable, that would require the appropriate actual authority review and a different disposition. Classifying the restriction explains which response route can be repaired without pretending that information itself will make every action available.

The baseline deserves equal scrutiny. A sponsor might describe the study as enabling Commit even though the institution would otherwise Stage, wait for another signal or take an already available protective response. The benefit must be assessed against that actual policy rather than an invented do-nothing baseline. Likewise, an action whose current expected value is positive may still be infeasible or dominated under an omitted consequence. The app can calculate a declared monetary account, but the decision owner must establish that the account includes the relevant option and consequence perimeter.

The restricted case is not a forecast that regulation or authority will remove Commit, and the complete case is not proof that Commit should be authorised. They are controlled alternatives. If a real restriction is unresolved, record conditional cases and the evidence needed to select among them. Avoid averaging those cases using invented probabilities to restore a reassuring headline. The useful result may be a hold on

commissioning until a response route is established, a smaller study tied to a feasible action or a separate proposal to change the operating or authority constraint.

Reconcile event, payer and timing before interpreting the result. Instrument acquisition belongs once in test cost; conditional response consequences belong in action payoffs. A resource needed to make Commit admissible cannot disappear from both because another budget holder owns it, nor be charged in test cost and again in every payoff. This ownership check complements the component account by testing omitted and duplicated response costs.

The completed packet also supplies a consistency test for simultaneous changes. If the population or reporting rule changes, LAB-A expires even when MENU-A is renewed. If only a cost allowance changes, the component account and testCost must be reconciled while unchanged likelihood settings can remain. If only response capacity changes, the admissible menu must be rebuilt while the accepted instrument account remains a separate object. A hypothetical revision that restores Commit and lowers price at once should retain both changes rather than attribute the whole improvement to renewed authority. This makes the next appraisal a controlled explanation and keeps every accepted record attached to the property it actually establishes.

Commission a usable branch policy

Completed original institutional packet for the fictional paired studies. The entries identify responsibilities; they assert no actual delegated authority or delivered research outcome.

Commission a usable branch policy		
Decision question	Packet entry	What closes the inquiry
What does the signal observe?	The same supplied binary instrument in both studies.	Relevant event and population evidence for its performance.
Which branch response changes?	Commit is removed; Stage and Hold remain.	A verified admissible menu from the response authority.
What remains unchanged?	Prior, likelihood settings, price and retained action payoffs.	Reproduction of both exact fixtures and branch outputs.
Who bears the consequence?	Identify the owners of operating and financial effects.	An agreed consequence perimeter and decision mandate.
What happens if response permission is missing?	Hold the full-menu commissioning conclusion.	Resolve admissibility or assess the feasible restricted policy.
What if the opportunity changes while research runs?	State that the current contrast contains no dated delay.	A suitable timing model and evidence where the decision requires it.

A commission also needs a timing account when the opportunity is transient. The reduced-upside study elsewhere in the paper changes a supplied payoff matrix; it does not estimate delay or assign a dated opportunity loss. A real timing question should identify commissioning duration, reporting latency, the response window and what evolves during that interval. Where the window governs the result, a dated model is needed. The annual or static expected-value comparison can still frame questions, but it cannot supply an undisclosed delay calculation.

Walk both adverse and favourable branches. The packet defers purchase because the priced, accepted instrument cannot deliver unrestricted response value; represented Stage and Hold still make the restricted comparison complete. Reopen only with prerequisites evidenced before the deadline and a canonical rerun of the current cost account. Proposed permission, incompatible likelihood support or superseded allowances fail that test. The branch walk audits the next fictional action without establishing actual legal authority, field likelihoods or institutional losses.

A different commissioning purpose needs a different mandate

An institution can reasonably commission evidence for obligations other than expected monetary gain: legal assurance, public accountability, health protection or learning for a later decision. Those purposes need their own mandate and evidence criteria. The present negative net information result does not prohibit them; it says that this declared static monetary policy does not justify the retained price under the restricted action menu. A separate purpose should be stated directly rather than hidden as an unmodelled benefit that supposedly repairs the numerical account.

The declared decision object governs transfer

The paired cases contain no actual institution, measured instrument or executed authority. Their arithmetic is a risk-neutral binary expected-value illustration. Feasibility, utility, correlated evidence, adaptive stopping and dated opportunity changes require additional methods where relevant. The packet is a reusable way to identify the missing decision object, not a certification that a particular research commission is appropriate.

Decision implications

- Information value depends on the admissible response menu as well as the instrument.
- Instrument assurance, consequence evidence and authority evidence answer different questions.
- A negative static monetary result does not adjudicate an unstated legal or public mandate.

PRIMARY EVIDENCE

References and source scope

jackson-voi

Christopher Jackson and coauthors; University of Cambridge repository. A guide to value of information methods for prioritising research in health impact modelling. *Epidemiologic Methods* 10(1), 20210012; November 2021

Locator: Sections 2 and 3; decision-theoretic expected loss and value of information

Scope: Authored tutorial in health-impact modelling. The paper here uses original illustrative commercial payoffs, not its empirical example or its broader algorithms.

<https://api.repository.cam.ac.uk/server/api/core/bitstreams/7da8efd7-7cb0-4424-963a-db7635e525cc/content>

efsa-methods

European Food Safety Authority. EFSA Tutorial on Uncertainty: Overview of methods for Uncertainty Analysis. Public tutorial, accessed October 2026

Locator: Methods for obtaining expert judgements; investigating influence and sensitivity

Scope: Food-safety assessment context. The inaccessible 2014 full guidance is not represented as read; this card cites the actual readable tutorial.

<https://multimedia.efsa.europa.eu/uncertainty-tutorial/principles-and-methods/methods.html>

magenta-methods

HM Treasury. Magenta Book Annex A: analytical methods for use within an evaluation. 15 May 2026

Locator: Analytical methods

Scope: Methods must match the available design and data. This paper does not imply that observational correlations establish causality.

<https://www.gov.uk/government/publications/the-magenta-book/magenta-book-annex-a-analytical-methods-for-use-within-an-evaluation-html>

magenta-book

HM Treasury. Magenta Book: Central Government guidance on evaluation. 15 May 2026

Locator: Evaluation planning; test and learn; benefits management

Scope: A monitoring improvement is not by itself a causal impact estimate. Identification and context remain necessary.

<https://www.gov.uk/government/publications/the-magenta-book/magenta-book-central-government-guidance-on-evaluation-html>

green-book

HM Treasury. The Green Book. 5 February 2026

Locator: Options appraisal; risk and uncertainty; monitoring and evaluation

Scope: Its social appraisal perspective and public discount conventions do not automatically apply to private finance.

<https://www.gov.uk/government/publications/the-green-book--2/the-green-book-2026>

orange-book

HM Treasury. The Orange Book: Management of Risk, Principles and Concepts. 29 July 2026

Locator: Part I principles; Part II risk control framework

Scope: UK central government requirements are described in their scope. They are not asserted as duties for every organisation.

<https://www.gov.uk/government/publications/orange-book/the-orange-book-management-of-risk-principles-and-concepts>

ipcc-uncertainty

IPCC Cross-Working Group Meeting. Guidance Note for Lead Authors of the IPCC Fifth Assessment Report on Consistent Treatment of Uncertainties. November 2010

Locator: PDF pages 2-4; evidence, agreement, confidence and likelihood

Scope: Climate-assessment guidance for AR5 authors; not a universal calibration scale or an industrial probability estimate. Supporting material did not undergo formal IPCC review.

<https://www.ipcc.ch/site/assets/uploads/2018/05/uncertainty-guidance-note.pdf>

nasa-se

NASA. NASA Systems Engineering Handbook. NASA/SP-2016-6105 Rev 2, 2016

Locator: Section 6.8 Decision Analysis; technical management processes

Scope: NASA engineering practice is a reference architecture. Transfer to a commercial organisation requires adaptation.

https://www.nasa.gov/wp-content/uploads/2018/09/nasa_systems_engineering_handbook_0.pdf

nasa-ridm

National Aeronautics and Space Administration. NASA Risk-Informed Decision Making Handbook. NASA/SP-2010-576, April 2010

Locator: Sections 2 and 3; performance objectives, alternatives, risk analysis and deliberation

Scope: NASA direction-setting guidance is adapted as an analytical architecture. Its mission context does not establish project-specific commercial preferences.

<https://ntrs.nasa.gov/api/citations/20100021361/downloads/20100021361.pdf>

nist-measurement

National Institute of Standards and Technology. Essentials of expressing measurement uncertainty: basic definitions. NIST Technical Note 1297 companion, accessed October 2026

Locator: Basic definitions; measurement equation; Type A and Type B evaluation

Scope: Metrology guidance; standard uncertainty is not a probability of commercial success or a general posterior model.

<https://physics.nist.gov/cuu/Uncertainty/basic.html>

nist-risk-framing

National Institute of Standards and Technology. Managing Information Security Risk: Organization, Mission, and Information System View. Special Publication 800-39, March 2011

Locator: Section 3.1, printed pages 33-36; risk framing

Scope: Information-security and US federal institutional context; adaptation does not confer compliance or prescribe a financial loss function.

<https://nvlpubs.nist.gov/nistpubs/legacy/sp/nistspecialpublication800-39.pdf>

aqua-book

UK Government Analysis Function. The AQUA Book. current edition accessed 2026-10-10

Locator: Quality assurance of analysis

Scope: A quality framework is not a certification of a particular model. Evidence of executed checks is still required.

<https://www.gov.uk/guidance/the-aqua-book>

uk-mca

UK Ministry of Housing, Communities and Local Government. Multi-criteria analysis: a manual. January 2009 amendment to 2000 manual

Locator: Weighting; sensitivity analysis

Scope: Preference weights express judgment. They are not event probabilities or observed causal effects.

<https://www.gov.uk/government/publications/multi-criteria-analysis-manual-for-making-government-policy>

sheffield-model-evidence

University of Sheffield, NICE Decision Support Unit. Reviewing model parameters: NICE Decision Support Unit technical support documents. Public TSD 13 publisher description, accessed October 2026

Locator: Publisher description of TSD 13, identifying and reviewing model evidence

Scope: Only the readable publisher description is used. The linked PDF was blocked in this retrieval and is not claimed read or paraphrased.

<https://www.sheffield.ac.uk/nice-dsu/tsds/reviewing-model-parameters>

doe-mv-plan

US Department of Energy, FEMP. Measurement and Verification Activities Required in the Energy Savings Performance Contract Process. current publication page accessed 2026-10-10

Locator: M&V plan and baseline conditions

Scope: A baseline must be defended for the actual project. The cited process does not validate illustrative inputs.

<https://www.energy.gov/cmei/femp/measurement-and-verification-activities-required-energy-savings-performance-contract>

doe-mv

US Department of Energy, FEMP. Measurement and Verification for Federal Energy Savings Performance Contracts. current publication page accessed 2026-10-10

Locator: Degree of measurement and verification

Scope: US federal contract requirements retain their stated scope; the analytical principle can be adapted without importing legal obligations.

<https://www.energy.gov/cmei/femp/measurement-and-verification-federal-energy-savings-performance-contracts>

fda-test-reporting

US Food and Drug Administration. Statistical Guidance on Reporting Results from Studies Evaluating Diagnostic Tests. Nonbinding guidance, March 2007

Locator: Sections 3, 5 and 7.1; definitions and reporting context

Scope: Medical diagnostic-device reporting guidance only. It does not validate the hypothetical commercial signal, establish clinical advice or supply its likelihoods.

<https://www.fda.gov/regulatory-information/search-fda-guidance-documents/statistical-guidance-reporting-results-studies-evaluating-diagnostic-tests-guidance-industry-and-fda>

gao-cost

US Government Accountability Office. Cost Estimating and Assessment Guide. GAO-20-195G, 12 March 2020

Locator: Technical baseline; assumptions; sensitivity; update with actuals

Scope: An earned-value metric cannot repair an invalid baseline or prove the quality of completed work.

<https://www.gao.gov/products/gao-20-195g>

gao-schedule

US Government Accountability Office. Schedule Assessment Guide: Best Practices for Project Schedules. GAO-16-89G, 22 December 2015

Locator: Ten schedule best practices; schedule risk

Scope: A deterministic critical path is a model result. It is not a probabilistic completion guarantee.

<https://www.gao.gov/products/gao-16-89g>

nist-risk-assessment

US National Institute of Standards and Technology. Guide for Conducting Risk Assessments. SP 800-30 Revision 1, September 2012

Locator: Prepare, conduct and maintain assessment

Scope: Information-system examples require adaptation outside that scope. Ordinal scales do not become quantitative probabilities through multiplication.

https://www.nist.gov/publications/guide-conducting-risk-assessments?pub_id=912091

PUBLICATION RECORD

Methods, limitations and review

This edition uses publicly accessible primary sources. All worked numerical examples are illustrative, with frozen inputs and reproducible calculations. They represent no actual client, mandate or operating project.

Publication fields identify the fixed edition dated 2026-10-10. Independent assessment and release verification are recorded against that fixed edition separately; these fields do not assign a quality score.

2026-10-10: Complete canonical manuscript, verified public source receipts and reproducible illustrative studies assembled. Mechanical publication checks support candidate readiness; independent review remains pending.

2026-10-10: Added explicitly fictional computational boundary studies and revised canonical model.1.0 precision policy. Independent assessment remains pending.

2026-10-10: Completed a fictional commissioning cost account and owner, evidence-state, expiry and present-disposition record in the final chapter; removed the repeated late comparison table. Independent acceptance remains pending.

2026-10-10-method-alignment: Aligned declared arithmetic, equality and display-order explanations with the current canonical implementation. Illustrative numerical inputs are unchanged; independent assessment remains pending.