



ICM
HOLDINGS INTERNATIONAL

RESEARCH & DECISION STUDIO

Choosing what deserves capital

A transparent preference model plus sensitivity is more defensible than one opaque ranking.

Portfolio prioritisation and strategic options

METHODS / EVIDENCE / WORKED STUDIES

ENGLISH EDITION / 2026-10-10

THE DECISION

Choosing what deserves capital

A capital committee cannot delegate its decision to a preference ranking without first defining which proposals are admissible, whose objectives govern the comparison and what commitment is being authorised. This methodological study develops an eligibility-first architecture for transparent scoring, range-aware weighting and decision-directed sensitivity. Its illustrative studies reproduce the Portfolio compass engine and expose both useful preference reversals and failures that a score alone cannot repair. The analysis distinguishes evidence about performance from judgement about tradeoffs, and individual option preference from the selection of a feasible funded programme. It supplies records for option development, evidence review, stakeholder disagreement, committee challenge and learning after commitment. The central claim is conditional: transparency improves the ability to challenge and revise a recommendation when the model represents a legitimate decision problem. It does not make an inappropriate aggregation valid or establish that the selected investment will succeed.

Questions this study answers

- What objective, eligible option set and commitment authority define the capital decision before candidates are scored?
- Which scales, tradeoff weights and performance evidence support the preference model, and which changes reverse the ranking?
- When do budget limits, dependencies or unacceptable performance require a different feasible programme from the ranked list?
- What evidence, challenge record and reopening event justify the next bounded commitment or require another decision method?

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 A ranking does not define the capital decision ..	4
02 Objectives must be distinguishable.....	9
03 The option universe is chosen before scoring ..	15
04 Evidence quality belongs beside performance..	21
05 Scales encode value judgments.....	26
06 Weights concern tradeoffs, not importance alone	31
07 Compensation can hide unacceptable performance.....	37
08 The baseline ranking is a conditional result	42
09 Sensitivity should seek preference reversals	48
10 Missing values are decisions in disguise.....	54
11 Uncertainty is not one confidence multiplier...	60
12 Budget selection differs from rank order.....	65
13 Dependencies determine strategic options	71
14 Different audiences can prefer different portfolios	76
15 A score can be gamed before the meeting	82
16 Committees need a challengeable recommendation	87
17 A pilot can buy decision flexibility	92
18 Performance review must preserve evaluation logic	98
19 When transparent scoring is the wrong instrument	103
20 The strongest case against preference models	109
21 Appendix: preference model specification	114
22 Appendix: worked illustrative portfolios	122
23 Appendix: facilitation and committee worksheets	131
24 Appendix: source scope and a dominated-option reversal	138

DECISION ARCHITECTURE

01

A ranking does not define the capital decision

A portfolio decision charter

Define the commitment before interpreting the rank

A committee presented with a ranked list is being asked to accept several decisions that may never appear on the first page. Someone chose which proposals would enter the list, which attributes would count, how performance would be measured and how much an improvement on one attribute could compensate for deterioration on another. A calculation then arranged the proposals according to those choices. The arithmetic may be impeccable while the decision remains badly specified. The useful opening question is therefore what the committee is authorised to choose, rather than which proposal has the highest displayed score.

This study treats a ranking as conditional evidence within a capital decision. Its value comes from revealing how supplied performance and declared preferences produce a recommendation. It can help a committee identify a promising option, locate a disputed assumption or explain why a change in priorities alters the preferred choice. It cannot, by itself, establish that an option is legally admissible, operationally feasible, affordable or compatible with other commitments. Those propositions require their own evidence and authority. A transparent ranking deserves attention only after its relationship to those requirements is explicit.

Consider a fictional operator comparing a reliability programme, an expansion proposal and a commercial learning initiative. The reliability programme may protect existing service, the expansion may increase capacity, and the learning initiative may improve the quality of future commitments. The alternatives do not necessarily perform the same function. Before scoring them together, the sponsor must explain the resource they compete for and the organisational objective that makes them comparable. If each addresses a different mandatory obligation, a single preferred option may be the wrong output. If they compete for discretionary capital within an agreed strategy, preference comparison can be useful.

The decision also needs a defined commitment. Approval to investigate a proposal differs from approval to acquire equipment, sign a delivery contract or release construction funding. A proposal that is attractive for further study may be unsuitable for immediate implementation. Treating those decisions as interchangeable rewards polished concepts over executable work, or suppresses potentially valuable ideas because their evidence is appropriately immature. A charter should identify the decision stage and the evidence required at that stage. The same candidate may receive different treatment as its scope, commitments and uncertainty become clearer.

A ranking hides timing particularly easily. A readily executable proposal can score well on readiness while an enabling investment creates benefits mainly through later options. If the committee is choosing an immediate intervention, readiness may be relevant. If it is establishing a long-term capability pathway, penalising every enabling proposal for delayed benefit may distort the strategy. Neither interpretation is inherently correct. The decision horizon and the treatment of dependencies should determine which comparison is appropriate. A proposal must not be made to resemble a standalone investment merely because the scoring form has no place for its role.

Evidence and scope

Public appraisal distinguishes objectives, options and uncertainty.

Evidence: [The Green Book](#)

The charter proposed here separates admissibility from preference. Admissibility concerns conditions an option must meet before the committee can consider a discretionary tradeoff. A legal prohibition, a binding technical interface or an unfunded minimum commitment cannot be repaired by a high score on strategic impact. Preference concerns differences that the authorised decision maker is willing to exchange. More initial expenditure might be accepted for better service, earlier delivery might be accepted with a narrower scope, or an option with uncertain benefit might be preferred when its learning value is important. The separation prevents an additive calculation from silently turning every requirement into a negotiable attribute.

The separation needs care because a supposed requirement can also conceal a preference. A sponsor may describe a preferred implementation date as mandatory when no external obligation makes it so. A technical team may reject an alternative for departing from a familiar architecture

even though the departure is feasible. Each admissibility condition should therefore identify its source, the authority that can interpret it and the evidence that demonstrates compliance. Where the authority permits an exception, the exception process belongs in the decision record. A gate that nobody can justify is no more legitimate than an unexplained weight.

Affordability is another distinct question. A low capital-cost weight does not make funding available, and a high cost weight does not impose a budget. The committee needs to know whether the candidate can be funded within the relevant cash and commitment limits. It also needs to know whether approving that candidate would foreclose another commitment. An initial comparison can leave the funding combination unresolved, provided the limitation is visible. Final capital authorisation cannot. A recommendation should state whether it identifies an individually attractive proposal, a feasible bundle or a sequence that still requires financial structuring.

Strategic fit should be tested with a traceable connection to an objective. A proposal is not strategically relevant merely because its description repeats the organisation's vocabulary. The sponsor should describe the mechanism linking the proposed intervention to an outcome, the assumption that makes the mechanism work and the evidence available to challenge it. A data platform, for example, may improve operating decisions only if information reaches people who can act and if those people have incentives to use it. The platform's technical completion is an implementation result. Its strategic benefit depends on a wider operating arrangement.

The committee should also specify its comparison universe before interpreting a winner. A list assembled from proposals that happened to arrive at the same meeting may omit a simpler way to achieve the objective. Improving an existing process, changing a contract or deferring commitment can be relevant alternatives to acquiring an asset. The analyst need not claim to have imagined every possible option. The analyst should explain the search process, inclusion rules and reason for closing the universe at the decision date. An admitted omission becomes a visible limitation; an invisible omission becomes an unearned claim that the ranking identifies the best available choice.

Separate the authorities behind a capital recommendation

The chair first identifies the commitment requested and the person authorised to approve it. The sponsor then distinguishes externally binding conditions, organisational requirements and discretionary preferences. The analyst documents the admissible option universe and performs the comparison within that universe. Finance confirms the funding and commitment perimeter. Delivery owners confirm the implementation dependencies. The committee can then approve, reject, stage or request evidence, recording which condition drives the disposition. A favourable score is supporting evidence within this chain of authority. It is not a substitute for the chain.

A charter that makes the missing decision visible

This original charter map is a diagnostic record, not an approval checklist that can establish compliance by itself. Read the final column as a reason to pause or redesign the comparison before accepting the score.

A charter that makes the missing decision visible		
Decision element	Record to establish	Failure revealed
Commitment requested	A defined scope and decision stage	A study recommendation is being presented as implementation approval.
Authorised objectives	Named outcomes and the authority choosing tradeoffs	Analysts or sponsors have substituted their preferences for the committee mandate.
Admissibility	Conditions, interpretation authority and evidence	Good performance elsewhere could compensate for a prohibited or infeasible feature.
Option universe	Inclusion rules and a documented search	The displayed winner has not been compared with a relevant alternative.
Resource perimeter	Funding, commitment and delivery limits	An individually attractive proposal cannot be funded or staffed with existing obligations.
Decision reopening	Material change triggers and review responsibility	The recommendation may remain in force after its decisive assumptions fail.

Original synthesis. The map separates calculation validity from the authority to make a commitment.

A useful charter is short enough to remain active during the meeting. It should not attempt to reproduce the entire business case. Its purpose is to expose the propositions that organise the analysis and the points at which

a different answer requires a different method. The supporting dossier can hold technical specifications, evidence records and financial schedules. The charter should direct a reader to those records and identify unresolved questions. If it can only be interpreted by the person who prepared the spreadsheet, it has failed as a governance instrument.

The charter should preserve the reason for an exception. A committee may knowingly approve a proposal that does not lead the preference ranking because it meets a strategic obligation, creates an enabling capability or can be implemented while the preferred alternative remains conditional. That is not necessarily an analytical failure. It becomes a failure when the dossier disguises the exception by changing weights until the chosen proposal appears to win. The stronger record states the ranking result, the independent reason for the disposition and the conditions under which the exception remains justified. This makes later review possible without forcing every legitimate consideration into the score.

The analyst should test the strongest alternative explanation before recommending the method. Perhaps the decision concerns an emergency response in which delay itself violates the objective. Perhaps every admissible option has the same service outcome, making a well specified cost comparison sufficient. Perhaps a proposal dominates its competitors on every relevant dimension, reducing the value of elaborate weighting. Transparent preference analysis is useful when consequential tradeoffs need to be expressed and challenged. It is burdensome when it reproduces a decision that a simpler legitimate rule can already make.

The thesis therefore concerns defensibility rather than automatic correctness. A transparent model allows another person to identify a preference assumption, replace it with an authorised alternative and observe the consequence. Sensitivity can show whether the recommendation survives an admissible change. Those properties improve challenge and accountability when the model is appropriate. They do not guarantee agreement, and they do not prove that the inputs represent future performance. An organisation can publish every equation in an invalid model. Visibility gives reviewers the opportunity to discover the defect; it does not remove the defect.

The opening decision record should end with a disposition that follows from the evidence available. If the comparison is legitimate but a decisive input is unsupported, the next action may be targeted investigation. If a requirement makes the preferred candidate inadmissible, the next action may be to remove it or seek an authorised exception. If funding combination is unresolved, the next action may be portfolio construction

rather than more scoring. If the choice is robust within the declared challenge set and the commitment is feasible, the committee can consider authorisation with monitoring conditions. The result should make the next responsible action clearer.

The case for a quick ranking

A sponsor may need a lightweight screen before spending substantial effort on every idea. Requiring a complete business case before any comparison would suppress exploration and consume scarce analytical capacity. The proposed architecture allows a proportionate screen, provided the commitment remains proportionate too. A preliminary ranking can identify candidates for further work when its incomplete evidence and provisional universe are explicit. The requirement becomes stronger as the organisation moves from investigation to an irreversible commitment. The repair is to distinguish decision stages and evidence standards, rather than demand final certainty at the start or allow an exploratory score to govern final approval.

Authority remains outside the arithmetic

This methodological study does not establish the legal or fiduciary authority of a particular committee, the applicability of a regulatory requirement or the availability of project funding. Its charter is an original analytical record that must be adapted to the organisational delegation and obligations. The illustrative model has no eligibility field, binding budget, interaction term or scheduling engine. Those matters need explicit treatment outside its calculation. A user who treats the highest score as a funded, admissible and deliverable portfolio would be asking the model to establish propositions it does not calculate.

Decision implications

- A preference ranking is evidence within a defined capital decision.
- Separate admissibility, preference, affordability and implementation authority.
- Record a justified exception directly instead of adjusting the model to conceal it.

DECISION ARCHITECTURE

02

Objectives must be distinguishable

A criteria architecture and overlap test

Separate what is valued from how it is measured and delivered

A criterion should tell the committee what difference between alternatives matters and why that difference belongs in the decision. A list of attractive words does not yet do this. Resilience, innovation, efficiency and sustainability can each refer to outcomes, implementation methods or organisational aspirations. If their meanings remain open, sponsors can

describe the same feature favourably under several headings and reviewers can interpret the same score differently. The criteria architecture needs to establish a stable question before anyone supplies a performance value.

An outcome describes a state the organisation values, such as more reliable service or better information for an operating decision. A means describes how a proposal may produce that state, such as a sensor network, an additional maintenance crew or a changed operating procedure.

Confusing the categories can reward an expensive means even when another intervention produces the valued outcome more effectively. An organisation may have a legitimate architectural requirement, but it should identify that requirement directly. Preference for a familiar implementation should not acquire the status of an outcome merely through the name of a criterion.

A proxy stands between an outcome and what can be measured. The number of functioning sensors may be easier to verify than the quality of the decisions they improve. The proxy can be useful when the relationship is specified and the limitation is understood. It becomes misleading when the committee treats verified sensor installation as verified operating benefit. The evidence record should identify the link from measured attribute to valued outcome, the conditions supporting that link and the circumstances in which it could fail. A proxy without a mechanism invites metric substitution.

Constraints serve another role. A system may need to connect to an existing interface, meet a safety requirement or operate within an approved site envelope. These conditions define what the committee can consider. Their evidence may resemble a performance measurement, but their interpretation is different. A preference score can distinguish better and worse performance within the admissible envelope. It should not permit a benefit elsewhere to compensate for a prohibited departure. Keeping a constraint outside the additive score can make the comparison simpler and the authority behind it clearer.

Implementation attributes describe the route to an outcome. Readiness, available skills and integration complexity can matter because they change the timing, resource demand or likelihood of delivering benefit. They should not automatically be scored as benefits in themselves. A ready proposal with little useful outcome may be easy to deliver and still deserve little discretionary capital. Conversely, an immature enabling proposal may be important to a strategy, although the immediate commitment should remain small and conditional. The architecture should identify why an implementation attribute is relevant to this decision stage.

Evidence and scope

Decision analysis links objectives, alternatives, criteria and documented rationale.

Evidence: [NASA Systems Engineering Handbook](#)

Double counting occurs when the aggregation rewards the same valued difference through overlapping criteria. Suppose a proposal receives a strategic-impact judgement partly because it reduces service interruptions, then receives a resilience judgement based on those same reductions. If both scores enter the sum, the committee may give interruption reduction more influence than it intended. The problem is not that the criteria are correlated. The problem is that their definitions allocate the same consequence twice without an explicit preference decision. A causal map can make that allocation visible before weights disguise it.

Correlation alone is an insufficient deletion rule. Service reliability and customer trust may move together while representing distinct outcomes that a committee legitimately values. Initial cost and implementation complexity may be associated without meaning the same thing. Removing one criterion because its values resemble another can erase a legitimate concern. The overlap test should ask whether improving a proposal on the underlying mechanism changes several criterion judgements for the same reason. If so, the committee should decide whether it intends that cumulative influence. Statistical resemblance is a prompt for interpretation, not a substitute for it.

The opposite error is to assume independence merely because the criteria have different names. A readiness assessment may already include procurement risk, operating skills and technical maturity. An execution-risk score may include those same attributes. Giving both substantial weight can impose an unintended penalty on an unfamiliar proposal. An analyst can preserve both criteria when their roles differ, for example by defining readiness as evidence of an executable plan and risk as consequences of residual failure after that plan is established. The distinction requires actual scoring definitions. A heading alone cannot establish it.

Strategic impact is especially vulnerable to becoming an omnibus criterion. If it includes revenue, cost savings, risk reduction, stakeholder benefit and delivery confidence, adding separate cost, risk and readiness dimensions can recreate those concerns within the total. The analyst should unpack the impact narrative and decide which components belong elsewhere. In the Portfolio compass, impact is a supplied judgement rather

than a measured common outcome. The user therefore needs an external definition that excludes whatever is intentionally scored under the other dimensions. The app cannot inspect the prose reasoning behind a supplied number.

A criteria tree can help, provided the hierarchy is substantive. At the top sit the outcomes that justify intervention. Beneath them sit attributes used to judge differences in those outcomes. An implementation branch can identify how delivery conditions affect the usefulness of a commitment, while an eligibility register holds requirements that cannot be traded. Every leaf should have a description, direction of preference and evidence rule. The tree need not become an elaborate diagram. Its purpose is to prevent a broad label and its detailed components from entering the aggregation as separate unacknowledged rewards.

Different statements need different decision treatment

This original classification tests a proposed criterion by its role. The examples describe a fictional service-modernisation discussion. They are design illustrations and make no empirical claim about an operator.

Different statements need different decision treatment		
Statement	Role	Appropriate treatment
Service remains available when a component fails	Valued outcome	Define the relevant service boundary and compare demonstrated or expressly assumed differences.
Install the proposed analytics platform	Means	Ask which outcome the platform changes and compare other ways of achieving it.
Installed instruments are producing usable readings	Proxy and implementation evidence	Record the mechanism connecting readings to a useful operating action.
Connect without violating a required safety interface	Constraint	Resolve compliance before discretionary preference comparison.
Operators can execute the change with available skills	Implementation attribute	Explain how the attribute affects this commitment and avoid repeating it inside an execution-risk score.

Read the treatment column before turning a statement into a score. Misclassification changes what the committee is allowed to exchange.

A useful practical test is a thought experiment rather than a new formula. Ask a scorer to imagine a proposal whose interruption exposure improves while every other relevant feature remains unchanged. Which criterion

scores would move, and what consequence justifies each movement? If the scorer changes several values for the same benefit, the tree needs review. If the scorer changes none, an important outcome may have no representation. The exercise reveals how the definitions work before actual proposals create defensive positions around them.

The test should be repeated with cost and implementation maturity. If a cheaper proposal gains under cost and strategic impact because impact includes financial efficiency, the cost preference may be represented twice. If improved readiness gains under readiness and reduces execution risk, the committee should explain whether it values earlier implementation separately from a reduced likelihood of failed delivery. These can be distinct concerns, but their separation must be deliberate. The architecture should make it possible to describe a plausible change that moves a criterion without automatically moving every neighbouring judgement.

Evidence definitions should be stable across alternatives. A proposal should not receive an impact judgement from a detailed operating study while a competitor receives a judgement from an untested sponsor narrative without that difference being visible. The architecture should identify the acceptable evidence and the treatment of weaker support. It can allow different information at an exploratory stage, but then the recommendation should carry the resulting limitation. A common performance scale does not create common evidential quality. The latter belongs beside the supplied value in the record.

The architecture also needs a retirement rule for criteria. A concern may be important during option development and cease to distinguish the admissible shortlist. If every eligible proposal meets the same interface requirement, retaining it as a scored criterion adds presentation without distinguishing performance. If a requirement changes into a genuine preference within the admissible envelope, the change should be documented. Criteria should follow the decision problem, but they should not be edited opportunistically after a favoured proposal performs poorly. Change control protects both relevance and legitimacy.

Completed overlap review for a fictional operating proposal

The completed record illustrates how a team can repair a criteria definition without pretending to have estimated a correlation or causal effect. Each row identifies a specific ambiguity and the decision about its treatment.

Completed overlap review for a fictional operating proposal			
Draft criterion	Overlap concern	Repair	Remaining evidence question
Strategic impact	Includes avoided outages that are also counted as resilience	Limit impact to the agreed strategic outcome; record interruption consequences in the designated reliability assessment	Which service changes matter to the strategy and through what mechanism?
Delivery readiness	Includes maturity already reflected in execution risk	Define readiness as evidence of an executable near-term plan	Does the plan show procurement, operating ownership and integration arrangements?
Execution risk	Repeats general immaturity rather than a failure mechanism	Assess residual consequences and exposure after the proposed delivery plan	What can still fail and which response is committed?
Capital cost	Acquisition scope differs across proposals	Require a common cost perimeter before normalisation	Which enabling and integration commitments are missing from the quoted scope?

Original design illustration. The repaired definition remains a proposal for the committee to authorise.

Authorise a criteria architecture before scoring

The sponsor states the valued outcomes and maps each proposed criterion to a specific decision question. Reviewers classify requirements, outcomes, proxies and implementation attributes. The team then runs the hypothetical-change overlap test and records repairs. The chair confirms which differences are compensable and which conditions remain gates. Only after those definitions are stable should alternatives receive performance judgements and preference weights. If the architecture cannot distinguish a valued consequence from the evidence supporting it, further scoring should wait for clarification.

The case for a broad strategic-impact judgement

A committee may prefer a broad impact judgement because a detailed criteria tree can become brittle and politically cumbersome. Expert reviewers may be able to recognise strategic relevance without separating every component. That approach can be proportionate for an exploratory discussion. Its limitation is that others cannot tell which consequences the judgement includes or how it overlaps with separate dimensions. The repair is a short explanatory rubric and a record of the decisive mechanism, rather than an exhaustive taxonomy. Broad judgement remains usable when its boundary is explicit and the recommendation does not imply a precision the rubric cannot support.

Distinct definitions do not establish statistical independence

The architecture is a semantic and governance test. It does not prove independence of outcomes, estimate correlations or validate the measurement quality of a performance scale. Interactions may remain even after definitions are clear. A technology intervention can change operating behaviour, costs and risk through the same causal pathway. The additive app does not estimate those interactions. The team must either justify the approximation for its purpose or retain an explicit external interaction analysis before committing capital.

Decision implications

- A criterion needs a stable question, preference direction and evidence rule.
- Overlap concerns mechanisms and definitions; correlation alone does not establish redundancy.
- Treat requirements, outcomes, proxies and implementation attributes according to their different roles.

DECISION ARCHITECTURE

03

The option universe is chosen before scoring

An ex-ante option register

Make nomination, exclusion and closure part of the decision record

The option universe determines what a preferred result can mean. A calculation can identify the highest score among the supplied candidates without establishing that an important alternative was considered. If the universe consists only of proposals promoted by teams with budget influence, the result may reward organisational access rather than better ways to achieve the objective. The committee needs a record of how alternatives entered the comparison and why the search stopped. A complete spreadsheet is not evidence of a complete search.

Option development should start from the problem and the required outcome. A request to replace an operating system may conceal a need for more reliable information, faster response or lower support burden. The organisation could address that need through a different system, a change in process, a service arrangement or a narrower intervention. Starting from the preferred solution makes those alternatives appear peripheral. Starting from the outcome gives them a fair opportunity to demonstrate whether they can meet the same need with a different commitment.

The search should distinguish an option from a variant. A supplier configuration, a delivery phase and an operating model can each be described as alternatives, although they answer different questions. Mixing them in a single list can give a large concept several chances to win while representing a competing approach only through an inconvenient variant. The analyst should define the level of comparison. If the decision concerns an operating approach, the shortlist should compare operating approaches. Supplier variants can be evaluated within the selected approach through a later, appropriately specified procurement decision.

A credible universe includes the reference action against which intervention is justified. Continuing current arrangements may carry deteriorating service, maintenance expenditure or exposure to an unmet obligation. Doing the minimum necessary to remain admissible may require an active programme rather than literal inaction. Those distinctions matter when costs and benefits are compared. The reference is not a free option with no consequences. Its scope should be described with the same discipline as the proposed changes, including what would happen without discretionary expansion.

The reference also establishes a causal question for benefits. A proposal should not receive credit for improvements that the organisation expects to obtain through an already committed intervention. Nor should its cost perimeter omit enabling work that would not otherwise occur. The analyst should identify what is incremental to the reference action and what remains common to every alternative. The Portfolio compass does not perform that incremental analysis. It receives supplied raw values. A consistent reference is therefore part of the evidence architecture preceding the calculation, rather than a result that the score can establish.

Evidence and scope

Public business-case guidance develops a longlist before choosing a shortlist.

Evidence: [Guidance on Developing Business Cases for Projects and Programmes](#)

Inclusion rules should be stated before the team observes a convenient winner. A rule may require a credible route to meeting the service objective, a defined commitment perimeter and enough evidence to permit a meaningful exploratory comparison. It should identify who can nominate an alternative and how an omission can be challenged. The rule need not admit every conceivable concept. It should distinguish exclusion because an option cannot address the problem from exclusion because its sponsor has not yet developed the evidence needed for this decision stage.

A concept with incomplete evidence may belong in a development register without entering the immediate commitment shortlist. That treatment preserves the idea while preventing an untested promise from competing as if it were executable. The record should state what evidence would make the concept eligible for reconsideration. Otherwise exclusion becomes permanent by administrative inertia. A sponsor should be able to distinguish a rejected mechanism from an incomplete dossier. The former calls for a different idea; the latter may justify targeted work if the potential decision consequence is material.

A feasibility screen should not disguise an outcome preference. Rejecting an option because it is unlikely to score well on strategic impact is different from rejecting it because it cannot meet a required service condition. The former anticipates the comparison and can eliminate a meaningful tradeoff. The latter applies an admissibility rule. The distinction is particularly important when a low-cost or staged intervention provides less benefit than a larger project but remains a legitimate response to the objective. The screen should retain it unless the required outcome truly makes it inadequate.

Early dominance can simplify the search, but it needs an evidentially fair comparison. An option that is no better on any relevant attribute and worse on at least one may not need elaborate preference analysis. The conclusion depends on common definitions, comparable evidence and the absence of an omitted distinguishing feature. A less mature proposal should not be declared dominated merely because its benefit has not yet been established. Unknown performance is a reason to investigate, bound or exclude under an explicit evidence policy. It is not automatically the worst possible performance.

The search record should preserve rejected alternatives and their reasons. Keeping them visible allows a later reader to distinguish a considered option from a forgotten one. It also supports a sensible reopening process. A previously infeasible service arrangement may become credible after a supplier develops relevant capacity. An expansion concept may cease to be viable after the organisation changes its strategy. The decision universe is

bounded at a date for a purpose, not frozen for the lifetime of the organisation. A material change in that boundary needs a new comparison record.

An original option-universe register

This completed register describes a fictional service-reliability decision. It illustrates inclusion and deferral reasons, without claiming that any actual organisation has these options or obligations.

An original option-universe register			
Proposed response	Comparison role	Initial disposition	Reason and reopening condition
Maintain current arrangements with required upkeep	Reference action	Include	Provides the common counterfactual; describe service consequences and required expenditure.
Redesign operating procedures and escalation	Distinct intervention mechanism	Include for investigation	Could address the outcome without acquiring a new system; operating evidence still needs review.
Acquire an integrated monitoring system	Distinct intervention mechanism	Include for investigation	Compare the complete enabling and operating scope, rather than equipment supply alone.
Contract for a managed monitoring service	Alternative operating model	Include for investigation	Needs a credible service boundary and a consistent treatment of retained operator responsibilities.
Expand unrelated production capacity	Different strategic problem	Exclude from this comparison	Can be evaluated under a separate capital mandate; it does not address the defined reliability outcome.
Autonomous intervention without a resolved safety interface	Incomplete admissibility case	Defer	Reconsider only after the responsible authority confirms an admissible operating design.

The register preserves an alternative even when it is deferred. That keeps evidence needs distinct from a judgement that the mechanism is unsuitable.

Changing the universe can change the arithmetic even when the original proposals remain unchanged. The canonical app uses the supplied candidates to set the observed ranges for normalisation. Adding a very costly proposal can compress the normalised cost differences between

existing candidates. Adding an option with weak readiness can change the meaning of every unchanged readiness value. A score increase after the universe expands does not establish improved performance. It may reflect a different scale. The worked set-extension study later makes that mechanism explicit through frozen inputs and outputs.

This dependence creates a governance reason to preserve a stable candidate set during a comparison session. A team can revise the set when an inclusion error is found, but the change should trigger a fresh interpretation of scores and weights. Earlier score levels should not be presented as directly comparable with the new levels. The raw evidence remains useful for continuity. The normalised scores belong to the stated set. Keeping both views allows the committee to understand whether a changed recommendation arises from new performance evidence, a changed universe or a revised preference.

The organisation should not insert a poor candidate merely to make preferred proposals look stronger. A dominated reference can alter observed ranges while contributing no credible choice. This is not evidence that range normalisation is always unusable. It is evidence that the model's scale is partly governed by admission decisions. The appropriate repair may be a tightly defined eligible universe, a separate experiment with fixed external anchors or a method that does not use candidate extremes. The app implements the observed-range method; alternative scaling architectures in this paper remain conceptual unless a separate verified model establishes them.

A strong search process invites a challenge to the mechanism, rather than asking reviewers only to propose a different supplier. An operator might ask whether the organisation could change the demand it needs to serve, share a resource with another activity or remove a bottleneck through a smaller intervention. These ideas can alter the strategic option space. The challenge should remain proportionate: a committee considering a near-term operational response cannot indefinitely defer action while pursuing speculative alternatives. The record should state the search effort justified by the scale, reversibility and uncertainty of the commitment.

Closing the universe requires a reason that can be reviewed. A fixed decision date, a material cost of delay or diminishing value from additional search can justify closure. Convenience alone is weak justification when an omitted mechanism could plausibly change the commitment. The analyst should record significant unresolved alternatives and the condition under which they would reopen the

decision. This does not make the recommendation indecisive. It gives the committee a controlled way to act now while recognising that the set of considered possibilities is an analytical boundary.

Close the universe for a defined purpose

The sponsor first defines the outcome and reference action. The team generates distinct intervention mechanisms before narrowing to detailed variants. Reviewers apply stated admissibility and evidence rules, preserving deferred concepts with reopening conditions. The chair then records the candidate set, comparison date and search limitations. Any later addition or removal is treated as a model change that may alter scales and preference interpretation. A recommendation can describe the best supplied admissible candidate; it should not claim to identify the best possible investment without a defensible search basis.

The case for accepting the proposals already available

An organisation may face a genuine time limit and have little capacity to develop alternatives from first principles. Comparing the proposals already available can then be a legitimate restricted decision. The restriction should appear in the mandate and recommendation. It is more honest to select the preferred available proposal within a bounded search than to imply that a convenient list is an exhaustive strategic universe. If the commitment is reversible, the committee can combine immediate action with a later search. If it is difficult to reverse, the value of considering a materially different mechanism deserves greater attention before closure.

A register cannot prove exhaustiveness

No finite option register establishes that every useful intervention has been imagined or that a deferred concept could be made feasible. The proposed record improves traceability of nomination, exclusion and closure. It does not measure the value of additional search or calculate a real option. The app compares the supplied candidates and has no model of nomination incentives, market availability or the future arrival of technology. Those boundaries remain part of the recommendation even when the scoring calculation reproduces perfectly.

Decision implications

- Develop alternatives from the outcome before narrowing to solution variants.
- Record why options are included, deferred or excluded and how the universe is closed.
- A changed candidate set can alter scales even when raw performance is unchanged.

DECISION ARCHITECTURE

04

Evidence quality belongs beside performance

An evidence confidence and acceptance worksheet

Preserve the proposition, perimeter and support beside every value

A performance value and the evidence supporting it answer different questions. The value states how an alternative is represented in the comparison. The evidence record explains why that representation should be considered credible for the decision being made. A highly attractive value supported only by a sponsor's assertion should not become equivalent to a less attractive value established through a relevant operating record merely because both occupy numeric cells. Common formatting makes them look comparable before their evidential basis has been examined.

The record should begin with the proposition the value is intended to support. An acquisition quote may support the price of equipment within its stated scope. It may not support installed capital cost, integration expenditure or ongoing operating burden. A successful technical demonstration may support the ability of a component to perform a task under controlled conditions. It may not establish the ability of the organisation to operate the full system. The committee needs the proposition and perimeter before deciding how much confidence to place in a supplied criterion value.

Evidence can be direct, inferred or expressly assumed. Direct evidence still has a scope: a measured result belongs to a particular operating condition, period and measurement process. An inference uses that evidence to represent another condition or a wider system. An assumption may be necessary when neither is available, particularly in early option development. Those categories do not create a universal hierarchy of usefulness. A directly measured but irrelevant observation can be less useful than a carefully explained inference. Their value depends on the decision question and the consequences of error.

A source citation is not enough to establish transfer. A supplier may present a successful deployment in a different operating environment. That record can demonstrate that the approach has been used, while

leaving local integration, staffing, data quality and service requirements unresolved. The transfer argument should identify which conditions are similar, which differ and why the differences do not defeat the proposed mechanism. Without that bridge, an external example becomes a persuasive image rather than evidence about the contemplated commitment. The dossier should preserve the limitation instead of letting the citation imply local validation.

Comparable evidence needs a common perimeter. If a reliability proposal is evaluated over a mature operating process while a commercial initiative is evaluated over an ambitious future scope, their impact judgements may be answering different questions. If one cost estimate includes integration and another excludes it, normalisation cannot make the values comparable. The analyst should reconcile scope, time horizon and included obligations before the score is calculated. A precise comparison of inconsistent estimates is still an inconsistent comparison. Its numerical neatness can make the problem harder to notice.

Evidence and scope

Analytical assurance distinguishes verification from fitness for the intended decision.

Evidence: [The AQUA Book](#)

Confidence should initially be represented beside performance, rather than compressed into it. A committee can then distinguish a proposal with strong performance and weak support from a proposal with modest performance and strong support. Multiplying a performance score by a general confidence judgement obscures that distinction. It can also imply an expected-value interpretation that the confidence judgement does not justify. If a team uses such an adjustment, it needs a clear model of what the multiplier means. The additive Portfolio compass does not supply that model.

The alternative proposed here is an evidence acceptance rule. For each criterion, state the minimum support required at the relevant decision stage, identify weaker support explicitly and define what the recommendation can claim while the gap remains. A preliminary comparison may allow an assumed readiness value when it authorises investigation. A final commitment may require an executable delivery plan and evidence of operating ownership. The rule should be known before the committee discovers which proposal benefits from lenient treatment. Otherwise evidence standards can be adjusted opportunistically alongside weights.

An evidence gap should lead to a decision-directed question. Asking for more information in general often produces another large document without resolving the choice. The analyst should identify which uncertain value could change the recommendation and what observation would distinguish the relevant alternatives. If the leader remains preferred under all declared plausible values for a weakly supported attribute, investigation may still be useful for implementation planning, but its immediate selection value is limited within that challenge set. If the choice changes, the gap becomes a potential condition of approval or a reason to stage commitment.

Completed evidence acceptance record

This original record uses a fictional monitoring-system proposal to distinguish an observed component result from an assumed organisational benefit. It provides an evidence question and a restricted decision use for each representation.

Completed evidence acceptance record			
Representation	Available support	Permitted decision use	Evidence still needed
Equipment acquisition scope	A supplier offer defines the supplied components	Compare the stated acquisition perimeter as an exploratory input	Integration, commissioning and retained operating costs on the same scope.
Usable operating information	A controlled demonstration produces technically valid readings	Treat the information mechanism as technically plausible under demonstrated conditions	Performance with local data, operating conditions and maintenance arrangements.
Organisational readiness	A sponsor identifies a prospective owner	Keep readiness provisional in a development comparison	Evidence of authority, skills, capacity and an executable delivery plan.
Strategic benefit	A narrative connects faster detection to improved service	Use an explicit assumed impact judgement with a visible limitation	A credible chain from detection to intervention and valued service change.
Execution-risk judgement	Reviewers identify unresolved integration mechanisms	Run declared alternative scenarios for the unresolved exposure	A tested mitigation design and evidence about the remaining failure mechanism.

The entries are design illustrations. They establish no actual costs, operating results or probabilities.

The worked implementation-evidence studies use expressly authored bounds, rather than pretending to estimate uncertainty from data. They change the readiness and execution-risk representation of a fictional

digital-twin proposal while keeping the other supplied assumptions fixed. The comparison asks whether those changes alter the preferred option. It does not assign a probability to either bound or claim that the bounds contain every possible outcome. Their usefulness comes from exposing what the model would do under stated alternatives. A real committee would need evidence to justify why those alternatives are appropriate challenges.

The evidence ledger should preserve disagreement about interpretation. An engineer and an operator may read the same demonstration differently because they are concerned with different failure mechanisms. An economist and a sponsor may disagree about whether a claimed impact is incremental to the reference action. Averaging their judgements can conceal the issue. The analyst should record the proposition each reviewer accepts, the proposition each rejects and the evidence that could reconcile the difference. A disagreement can be a useful specification of the next investigation rather than noise to be removed.

The ledger also needs provenance that survives a meeting. A value should link to the record used, its date, the scope it covers and the person responsible for interpreting it. If a committee revises a value during discussion, the reason should be attached to the revision. A spreadsheet cell that changes without a record is difficult to challenge later. The analyst should be able to reconstruct whether the change reflected new evidence, a new interpretation, a corrected scope or a sponsor's preference. Those are different causes with different implications for review.

Estimates should not acquire false certainty through successive copying. A preliminary assumption can pass from a concept note to a business case and then to a presentation, gaining the appearance of institutional acceptance without gaining evidence. The provenance record should retain the original epistemic status until a documented review changes it. Approval of a document is not necessarily validation of every estimate it contains. The committee should know which values were checked, which were accepted provisionally and which were outside the review's scope.

A missing value should remain missing until a responsible treatment is chosen. Substituting a neutral score can favour an option by hiding a weakly supported attribute. Substituting the worst score can eliminate an option without evidence that its performance is actually poor. The choice between exclusion, bounded scenarios and information collection belongs to the decision policy. The canonical app rejects incomplete numeric input and does not infer a missing value. That behavior makes the unresolved policy visible instead of performing an unexplained imputation behind the interface.

Assurance should be proportionate to consequence. A reversible trial can proceed with less complete evidence than a commitment that creates a long-lived dependency, provided the trial is designed to resolve a specific question. A minor uncertainty about a non-decisive attribute may not justify the same work as an unsupported assumption governing the preferred option. Proportionality does not mean lowering standards until a proposal passes. It means directing scarce verification effort toward the propositions that could change the decision or leave the organisation unable to respond after commitment.

The evidence discussion should end with a controlled use statement. The analyst can say that a value is supplied for illustration, accepted for screening, bounded for a challenge or established for a defined operational scope. The committee can then attach a condition to its recommendation. This is more informative than labelling an entire model high confidence while leaving individual assumptions unexplained. The record should make clear what the evidence permits the organisation to do now and what remains to be established before the next commitment.

Direct assurance toward the proposition that governs the choice

The analyst maps each criterion value to its proposition, perimeter and retained supporting record. Reviewers identify unresolved transfer, scope or interpretation questions. The team runs declared alternative representations for values that could change the recommendation. The chair then chooses whether to accept the evidence for this stage, restrict the commitment, require a specific observation or remove the option from the current shortlist. Every revised value receives a reason and provenance. Assurance is complete only for its stated use; no general confidence badge substitutes for that record.

The case for a single confidence adjustment

A single confidence factor is attractive because it appears to penalise optimistic claims consistently and keeps the result simple enough for a committee to scan. It can be appropriate when a validated model gives the factor a precise interpretation. In an ordinary preference workshop, confidence often combines evidence quality, delivery concern and personal belief. Multiplying by it can repeat risk already represented elsewhere and conceal why a proposal loses value. The repair is to show performance and support separately, then use targeted scenarios or an explicitly justified adjustment. Simplicity is useful when the meaning remains defensible.

An evidence ledger is not a validation certificate

The proposed ledger documents the basis and permissible use of supplied values. It does not prove that a supplier record is accurate, that a demonstration transfers to another setting or that reviewers have interpreted it correctly. Those questions need appropriate technical, commercial and operational assurance. The illustrative bounds are not calibrated probability distributions. The paper and app calculate no confidence level for the preferred option. A reproducible score shows the consequence of the supplied representation, rather than the truth of the represented future performance.

Decision implications

- Common numeric formatting does not establish common evidence quality.
- Record what each source supports and what transfer or scope argument remains unresolved.
- Use decision-directed assurance and explicit scenarios before compressing confidence into a score.

MECHANISMS AND WORKED STUDIES

05 Scales encode value judgments

A scale-choice sensitivity experiment

Interpret the yardstick before interpreting the gap

A common scale makes attributes arithmetically comparable by choosing what differences mean. It does not discover that meaning in the data. Capital expenditure, a strategic-impact judgement and a readiness judgement arrive in different units and with different interpretations. Turning them into a common range allows addition, but the transformation determines how much an observed difference contributes to the total. The scale choice therefore belongs in the preference architecture. It should be explained before the committee interprets a score gap as evidence of a better commitment.

The canonical app uses an observed-range transformation. For an attribute to be maximised, the least favourable supplied value receives the low end of the common scale and the most favourable supplied value receives the high end. Intermediate values are placed linearly between those extremes. Cost and execution risk reverse the direction because lower supplied values are preferred. When every candidate has the same value on an attribute, the app assigns a common midpoint. This is a specified calculation convention. It makes no claim that the candidate extremes represent an absolute failure or ideal outcome.

The convention has a useful property for exploratory comparison. It preserves the direction of raw preference and displays how each alternative sits relative to the available set. It also prevents currency amounts from overwhelming a dimension simply because their units are numerically larger than an index. Those benefits are conditional on the raw values being comparable and the linear transformation being an acceptable representation of value. A transformation cannot repair inconsistent cost scopes or establish that equal changes on an ordinal judgement carry equal preference value.

Linearity is a substantive assumption. The organisation might value the first improvement in reliability more than a later improvement after service is already acceptable. It might place little value on readiness below an executable threshold and substantial value on reaching that threshold. An impact rubric may describe qualitative categories whose numerical spacing is only a convenience. The app does not estimate a nonlinear value function or infer these thresholds. A user who adopts its linear transformation should understand what the approximation permits and where it may distort the intended tradeoff.

Observed ranges also give the candidate universe control over the scale. If the best and worst supplied values move, the same raw difference can represent a different portion of the normalised range. This is not an implementation error. It is how the chosen transformation works. It becomes a decision error when the organisation compares scores from different sets as if their scale had remained fixed, or retains weights elicited for a different swing in performance. The interpretation of the weight and the interpretation of the scale must remain connected.

Evidence and scope

Normalisation, weighting and sensitivity are linked choices in composite-indicator construction.

Evidence: [Handbook on Constructing Composite Indicators: Methodology and User Guide](#)

The set-extension experiment is designed to isolate that dependence. Its starting point is the balanced illustrative mandate. The extension retains every existing alternative, raw criterion value and declared weight, then adds an expressly inferior reference proposal. The added proposal has weaker impact and readiness, greater cost and worse execution risk than the existing candidates. The experiment is not an argument for admitting such a proposal. It asks what the canonical observed-range engine does if the universe contains it, so that the governance consequence becomes visible.

The important comparison is between the unchanged alternatives, rather than between the new proposal and a winner. If their raw performance and weights stay fixed while their preference scores change, the change must be interpreted through the new scale. The experiment therefore retains both the input record and the computed output. It does not describe a higher score as improved operating performance. It does not describe the added proposal as a newly available market option. Every supplied quantity remains explicitly illustrative and belongs only to the frozen study.

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

Illustrative scores change when the candidate set changes

The existing raw values and declared weights are held constant. Only a dominated reference proposal is added. The score differences are consequences of the current-set range transformation, not new performance evidence. Read each score within its named universe.

Illustrative scores change when the candidate set changes			
Unchanged alternative	Supplied capital cost	Balanced-set score	Extended-set score
Reliability architecture	24.00	72.13	89.98
Capacity extension	90.00	35.00	78.23
Commercial learning	18.00	62.86	86.03
Energy productivity	42.00	59.32	85.73

Canonical Portfolio Compass calculations. Input and complete output are frozen in balanced-baseline and range-artifact. Capital cost is in USDm; preference points have no monetary or probabilistic meaning.

An illustrative range effect without a leader reversal

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

The balanced illustrative set gives Reliability architecture a score of 72.13 preference points. With the dominated proposal included, the same alternative receives 89.98 points and remains the displayed leader. The added proposal receives 0.00 points. This experiment demonstrates changed score levels while explicitly disclosing the absence of a leader reversal in these supplied cases. It cannot establish how frequently set changes would reverse a recommendation in any population of investment decisions.

The result changes the interpretation of a presentation. A sponsor who reports only the extended-set score might make an unchanged proposal appear more compelling than it looked under the original set. The higher number would be correct for the new calculation, yet misleading as evidence of improvement. The repair is to show raw values, name the universe and explain the transformation. A committee should be able to see whether it is discussing a better intervention or a different yardstick. The score cannot supply that distinction without the input record.

Fixed external anchors are a conceptual alternative. The organisation could define a relevant worst and best performance level independently of the candidates, then place each candidate against those anchors. That approach can improve continuity across candidate sets, but it creates other responsibilities. The anchors need a legitimate objective basis, an evidence perimeter and a review rule when the operating context changes. Anchors chosen merely to favour a proposal are no more defensible than a manipulated observed range. This paper describes the alternative architecture; the current app does not implement it.

A value function adds another distinction. It represents the committee's preference over performance levels, rather than merely rescaling an observed range. A function can express saturation, a threshold or an asymmetric concern if those features belong to the authorised preferences. It should not be fitted casually to make an inconvenient candidate lose. Eliciting and validating such a function requires a separate method and record. The canonical linear app is useful partly because its simplicity exposes where a richer preference model would be needed. It should not be presented as if it had already estimated that richer model.

A scaling review can therefore be a small set of decision questions. What raw difference does the full normalised swing represent? Is an equal increment in the raw value an acceptable equal increment in preference? Would adding or removing an irrelevant candidate materially change the interpretation? Are the supplied weights intended for this specific swing? Does a required threshold belong outside the compensatory score? The answers determine whether the observed-range convention is appropriate for the current purpose, rather than whether the mathematics can execute it.

The organisation should retain a scale version with the decision record. If the candidate universe changes, the version should change even when the raw source files for the existing alternatives do not. If a criterion definition changes, the team should revisit the raw values as well as the transformation. A later reviewer can then distinguish an updated operating assumption from a changed value convention. This discipline supports learning without treating every historical score as though it belonged to a single permanent scale.

Accept a scale for its declared use

The analyst first reconciles the raw evidence perimeter and preference direction. The committee identifies the performance swing represented by the chosen transformation and decides whether linear placement is acceptable. The team tests a relevant set or anchor change, preserving raw values so that altered yardsticks remain visible. Any required threshold is treated as an eligibility condition or a separately justified value-function feature. The recommendation names the scale and candidate universe. Scores from different versions are not treated as direct evidence of performance improvement.

The case for the observed range

Observed ranges are attractive when an exploratory team lacks legitimate external anchors. They use the candidate evidence available, are easy to reproduce and make the contrast within a shortlist clear. Demanding a sophisticated value function can consume effort without improving a preliminary decision. The appropriate response is not to reject range normalisation in every setting. It is to limit the interpretation to the stated set, connect weights to the represented swings and disclose sensitivity to plausible changes. The simple method remains useful when its users understand the yardstick they have chosen.

The experiment establishes a mechanism, not its prevalence

The dominated-set extension is deliberately constructed to expose a property of the algorithm. It is not a sample of real nomination practices, an estimate of manipulation frequency or a measured loss from poor capital allocation. The experiment also does not implement anchored scales or nonlinear value functions. Those approaches remain conceptual alternatives requiring their own specified and verified calculation before numerical results could be claimed. The canonical study supports only the stated consequence for its frozen candidate sets.

Decision implications

- Normalisation chooses how raw differences enter preference aggregation.
- Observed-range scores belong to the named candidate set and scale version.
- A higher score after a set change need not indicate improved performance.

MECHANISMS AND WORKED STUDIES

06

Weights concern tradeoffs, not importance alone

A facilitated weighting protocol

Give every weight a range, rationale and authorised use

A question about which objective is important is not yet a question about the tradeoff represented by a weight. An organisation may regard safe operation as indispensable while every admissible alternative satisfies the same required condition. The importance of safety then belongs in eligibility and governance, rather than necessarily distinguishing the scored options. Conversely, a modest-looking attribute can carry a consequential difference across the candidate set. A useful weight concerns the value of the performance swing being compared, not the rhetorical prominence of the criterion name.

Swing weighting makes that distinction explicit. The facilitator describes a relevant movement from the least favourable to the most favourable performance on a criterion, while holding the other represented attributes at a specified reference. The decision maker compares the value of that movement with the swings on other criteria. Relative weights express those comparisons. The question is not how much the organisation cares about the word readiness. It is how much it values the particular improvement in readiness represented by the chosen scale, relative to the other improvements the decision can buy.

The performance ranges therefore belong in the weighting discussion. If the cost swing represents a small difference between already affordable proposals, its preference value may be limited. If it represents a difference that consumes much of the discretionary capital envelope, the same criterion name can describe a more consequential swing. A weight elicited for one range should not automatically carry over to another. Observed-

range normalisation makes this especially important because the candidate universe determines the endpoints. The set and scale should be stable enough to support the elicitation being attempted.

Holding other attributes fixed is a thought experiment with a purpose, not a claim that every reference profile is physically feasible. It isolates the preference comparison. If the decision maker cannot value an improvement without knowing how another attribute performs, the interaction may matter. For example, improved monitoring information may be useful only when an operating team can act on it. A simple additive model can misrepresent that dependence. The facilitator should record the difficulty instead of forcing a clean weight from a question whose premise the participant rejects.

The relevant condition is about preferences, rather than statistical independence. Criteria can move together in reality while the organisation can still assess the value of their changes separately. Conversely, technically independent features can have joint preference value because one is useful only with the other. Clear definitions and an overlap review help, but they do not establish additivity by themselves. The committee should understand what kind of separability it is accepting when it uses a weighted sum. The app performs the sum; it does not test the committee's underlying preference structure.

Evidence and scope

Swing weights concern preference changes over defined criterion ranges.

Evidence: **Multi-criteria analysis: a manual**

A facilitation session should begin with the decision mandate and the scoring definitions. Participants should know whether they are advising an authorised committee, expressing different stakeholder perspectives or agreeing a collective preference record. Those activities can produce different legitimate outputs. A facilitator should not turn disagreement into a single weight vector merely because the software expects one vector at a time. The model can calculate each declared scenario separately, while the governance process decides which scenario, compromise or further discussion has authority.

The facilitator then presents the swings in concrete terms. Raw units and descriptions matter more than a generic common-scale endpoint. Participants should be able to explain what improves, what remains unchanged and why the improvement affects the objective. If they cannot interpret a swing, the criterion rubric or evidence perimeter needs work.

Assigning an arbitrary importance rating at that point creates a number without resolving the ambiguity. The later precision of the score will conceal, rather than correct, the unresolved interpretation.

A reference swing can organise the discussion. Participants identify a particularly valuable movement and express the remaining movements relative to it, checking the comparisons for consistency. The resulting raw weights are normalised by the engine so that a common change of units in the weight vector does not alter the aggregation. Normalisation removes an arbitrary common scale; it does not validate the relative judgements. A weight vector can be internally well formed and still reflect an inappropriate mandate or misunderstood performance ranges.

The record should capture reasons, not only values. A preference for readiness may reflect a near-term service obligation, an unwillingness to commit before operating ownership is established or a need to preserve scarce delivery capacity. Those reasons have different implications when the context changes. The committee should be able to identify which reason justified the weight and what event would require reconsideration. A vector copied into later meetings without its rationale can become an accidental policy that nobody remembers authorising.

The facilitator should also check the meaning of a zero weight. It can express indifference within the relevant admissible range, or it can reveal that a concern has been handled as a gate elsewhere. It should not silently mean that the concern is unimportant in every context. In the dominance study, an impact-only vector leaves candidates tied on weighted preference even though their raw cost, readiness and risk differ. That result is an opportunity to test the intended indifference. It is not a reason for the engine to invent a preference the user did not supply.

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

Declared illustrative weights across different mandates

The alternatives and every raw performance value remain identical across these frozen mandate studies. Only the supplied relative weights change. The values are illustrative judgements, not measurements of stakeholder opinion or frequencies in a population.

Declared illustrative weights across different mandates			
Criterion	Balanced mandate	Impact-led mandate	Initial-capital preference
Capital cost	25.00	5.00	75.00
Strategic impact	35.00	75.00	10.00
Readiness	25.00	10.00	10.00
Execution risk	15.00	10.00	5.00

Canonical fixture inputs for balanced-baseline, impact-mandate and liquidity-mandate. The engine normalises each supplied vector before calculating preference scores.

The illustrative mandates are deliberately different. They allow the paper to test how an authorised change in priorities can alter preference without pretending that a proposal's performance has improved. An actual workshop would need an accountable process for adopting its vector. The committee might choose the balanced mandate, give priority to a particular strategic outcome or impose an initial-capital preference. The calculation can reveal the consequence of each choice. It cannot determine which mandate the organisation is morally, legally or strategically entitled to impose.

Sensitivity belongs after the initial elicitation as a challenge to interpretation. A participant may discover that a proposed weight produces a recommendation inconsistent with a stated tradeoff. That discovery can justify revisiting the weight or criterion definition, provided the reason is documented. It should not become an invitation to tune the model until a favoured proposal wins. The difference lies in the rationale. Correcting a misunderstood swing improves the preference record. Changing a weight solely to secure a desired outcome conceals the real decision.

The team should examine joint changes as well as the app's individual shifts when the decision requires it. The canonical sensitivity routine moves a single normalised weight and rescales the others. It provides a transparent local challenge set, not an exhaustive exploration of every legitimate preference vector. A coherent stakeholder mandate may change several weights together. The frozen mandate studies provide such

comparisons through complete alternative vectors. Neither the individual tests nor the selected vectors establish an empirical distribution of preferences.

A facilitator should preserve dissent when it remains substantive. The record can state that a delivery owner places greater value on an executable near-term intervention while a strategy sponsor places greater value on a capacity outcome. Both may understand the criteria correctly. Their disagreement concerns priorities or authority, rather than an error in arithmetic. The committee can decide whether to adopt one mandate, seek a negotiated compromise or stage investments under different decision envelopes. An unexplained average hides the conflict while making the result appear consensual.

A completed facilitation record

This original qualitative record describes how a fictional committee could interpret a readiness swing. It contains no invented survey result or numerical elicitation. The purpose is to make the reason for a future weight inspectable.

A completed facilitation record		
Prompt	Recorded response	Facilitator check
What changes across the readiness swing?	The proposal moves from an unresolved operating arrangement to a plan with accountable owners and available delivery capacity.	Do the endpoints describe the same scope and commitment stage?
Why does that movement matter here?	The requested commitment concerns a near-term service intervention that cannot wait for an undefined operating model.	Is urgency a genuine decision condition or an unsupported sponsor preference?
What stays fixed in the comparison?	The stated strategic outcome and cost perimeter remain unchanged in the elicitation thought experiment.	Can participants value the swing without assuming that impact or cost automatically changes?
What would reopen the judgement?	A different commitment stage or a material change in the candidate readiness range.	Will the reason and range travel with the stored weight vector?

Original design illustration. An actual committee would supply and authorise its own tradeoff judgement.

The session should finish with a weight-use statement. It identifies the decision stage, authorised perspective, criterion definitions, ranges and rationale governing the vector. It also names a change that would reopen the elicitation, such as a different candidate set, a revised cost perimeter or a material change in objectives. That statement gives the weights a

controlled life. The organisation can then reuse the record when appropriate without treating a context-specific tradeoff as a universal preference policy.

Elicit a tradeoff the committee can explain

The facilitator confirms the mandate and criterion definitions, describes the raw performance swings and tests whether participants can value them with the other represented attributes held at the reference. Participants state relative preferences and the reasons behind them. The analyst records disagreement as separate scenarios where needed, then runs sensitivity to challenge the interpretation. The committee authorises the vector for a stated use and reopening condition. If the elicitation requires interactions the additive app cannot represent, the recommendation should identify that method boundary rather than force a convenient vector.

The case for direct importance ratings

Direct ratings are easier to explain than swing-weight elicitation and may be sufficient for a preliminary screen. Participants often understand that they are expressing priorities rather than estimating technical quantities. Their weakness is that the ratings may ignore the ranges the normalisation actually compares. The repair is a short range check: show what the full movement on each criterion represents and ask whether the relative ratings still express the intended tradeoff. When the decision is consequential or the ranges differ sharply, a more disciplined elicitation is justified. The convenience of the first question should not determine the validity of the final commitment.

No elicitation process is performed by the app

The model normalises supplied weights and calculates their consequences. It does not interview participants, establish authority, estimate interpersonal preferences or test additive separability. The illustrative vectors are authored scenarios. Their results cannot be used as evidence that actual stakeholders would choose those weights. A decision process must supply the mandate, range interpretation and rationale externally, with appropriate facilitation where the institutional framework requires it. Reproducible aggregation is a property of the calculation, not a certification of the preference elicitation.

Decision implications

- A weight concerns the preference value of a defined performance swing.
- Normalisation does not validate the relative judgements or decision authority.
- Preserve substantive disagreement as explicit scenarios before adopting a collective vector.

MECHANISMS AND WORKED STUDIES

07

Compensation can hide unacceptable performance

An eligibility-first decision framework

Resolve non-compensable conditions before interpreting the sum

Some conditions define whether an alternative may be considered at all. They do not describe a benefit that a sufficiently strong performance elsewhere can offset. A committee may require a lawful operating arrangement, an acceptable environmental and social assessment, an available delivery owner or a minimum service capability. An attractive strategic story does not necessarily compensate for failure on such a condition. The decision process should identify these requirements before aggregation, with authority and evidence appropriate to the condition being imposed.

Eligibility and preference answer different questions. Eligibility asks whether the option belongs in the admissible choice set at the specified commitment stage. Preference asks which admissible option better serves the stated objectives. A proposal can be eligible and undesirable, or potentially desirable and currently ineligible. Confusing these states makes remediation difficult. The team may need to redesign an attractive but inadmissible proposal, while an eligible weak proposal may simply deserve no further attention. The recommendation should state which problem exists rather than treating every low score as the same diagnosis.

A gate needs an accountable owner. The analyst should identify who may define the condition, who can verify its satisfaction and who can authorise an exception if exceptions are permitted. An invented numerical threshold is not a substitute for that governance. A risk rating in the app is an illustrative preference attribute; it does not establish compliance with an applicable standard or permit. A screening team must use the relevant assessment process and retain the evidence supporting its conclusion. The weighted model can then compare options that have passed the required checks.

A completed eligibility record

This fictional record separates permission to investigate from permission to implement. Its categories are procedural examples, not legal determinations or findings about a project.

A completed eligibility record			
Condition	Present commitment	Evidence owner	Disposition
Operating accountability	Authorise a bounded design study	Service sponsor confirms who owns the proposed service	Study may proceed with a named operating-design deliverable; implementation remains subject to approval
Environmental and social assessment	Develop an option comparison	Assessment lead identifies applicable process and information needs	Include the proposal as a development option; do not treat a favourable preference score as a completed assessment
Delivery capacity	Reserve implementation resources	Delivery director confirms available capability and competing obligations	Defer resource reservation until the capacity check is resolved
Strategic objective	Choose an intervention to advance the declared outcome	Committee confirms the objective and decision envelope	Exclude proposals that serve a different objective unless the envelope is formally reopened

Original completed illustration. The committee must replace the proposed checks with the requirements governing its actual decision.

The condition also needs a timing statement. A feasibility-study commitment may reasonably proceed while some operating requirements remain unresolved, provided the study is designed to resolve them and does not itself require those conditions to be satisfied. A construction commitment may require a different evidence standard. Calling a proposal ineligible without naming the decision stage can prematurely exclude a useful development option. Calling it eligible for a study does not authorise implementation. The register should distinguish the present commitment from the future project state.

The committee should avoid hiding a gate inside a very large weight. A high risk weight still participates in a compensatory sum. Enough benefit on other represented attributes can offset poor risk performance, depending on the scales and the candidate set. If a condition is truly non-compensable, represent it outside that sum. This preserves the meaning of

both mechanisms. The gate records the required condition; the weights record tradeoffs among admissible performances. A large weight can express strong preference, but it should not masquerade as a prohibition.

Evidence and scope

Environmental and social management has its own assessment and management requirements.

Evidence: Performance Standard 1: Assessment and Management of Environmental and Social Risks and Impacts

Dominance is a separate useful challenge. In the canonical app, a candidate is Pareto-dominated when another candidate is no worse on every represented raw criterion and strictly better on at least one, taking the declared directions into account. Lower cost and risk are favourable; higher impact and readiness are favourable. This test does not use the weights. It identifies an option that lacks a represented performance advantage over a comparator. Before retaining that option, the committee should ask whether an omitted attribute, different scope or data error explains the apparent disadvantage.

A dominance result depends on the represented attributes and their evidence. It cannot prove that an option is inferior on every consideration that matters to the organisation. A proposal may possess an unrepresented contractual right, geographic fit or strategic option. The proper response is to examine whether that feature belongs in the decision architecture, rather than dismissing the test or inventing an unexplained override. If the feature matters, document it explicitly. If it does not, the dominated option usually needs a reason to consume further analytical attention.

Dominance and an additive ranking can disagree in an instructive way when a criterion receives no weight. The weighted preference then deliberately ignores differences on that criterion. A candidate may tie on every positively weighted attribute while being worse on ignored attributes. The app reports the raw dominance result alongside the weighted tie. It does not silently modify the declared weights. This makes the tension visible and lets the committee decide whether its stated indifference is intentional, whether the ignored attributes belong in gates or whether the vector needs reconsideration.

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

A weighted tie with a raw dominance warning

This illustrative fixture assigns preference only to strategic impact. Both candidates have equal impact, while the cheaper proven system is better on the other represented raw attributes. The literal text display order of tied scores is not a substantive recommendation.

A weighted tie with a raw dominance warning						
Candidate	Capital cost	Impact	Readiness	Risk	Score	
A costly immature system	90.00	80.00	40.00	40.00	40.00	50.00
B cheaper proven system	50.00	80.00	90.00	10.00	50.00	50.00

Canonical ignored-dominance input and output. Capital cost is illustrative USDm; impact, readiness and risk are supplied ratings. Every printed value is bound to the frozen engine result or supplied illustrative input.

Interpret the disagreement

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

In this illustrative calculation the weighted scores are 50.00 and 50.00. The raw Pareto set contains B cheaper proven system. The result does not license the literal text display leader to receive capital. It asks the committee whether indifference to cost, readiness and risk is truly authorised, or whether those concerns are meant to be handled elsewhere. Equal weighted scores and unequal raw performance can both be correct statements under the declared model.

A useful review sequence therefore begins with eligibility, proceeds to a dominance challenge and then examines preference. The sequence does not require excluding every dominated proposal automatically. It requires an explicit explanation for keeping one. A development proposal may be retained because an uncertain attribute could change after targeted investigation, because its scope is being redesigned or because the represented comparison omits a material feature. Each reason should identify the work needed to resolve it. Retention without a reason weakens the discipline of the option set.

Gates themselves can be gamed. A sponsor might define a technical requirement so narrowly that only its preferred design satisfies it, or declare a delivery feature indispensable without showing how it serves the objective. The gate review should test necessity, proportionality and decision authority. Where different technologies can achieve the same required outcome, an outcome-based condition may preserve more

legitimate alternatives than a design-specific condition. This is a governance judgement informed by technical evidence, not a task the scoring app can settle automatically.

Exceptions deserve visible treatment. If the organisation has an authorised exception process, the committee should record the condition, rationale, compensating measures and person responsible for the residual decision. An exception should not appear as a favourable score adjustment. That would conceal both the departure and the authority that accepted it. Where an exception is unavailable, the analyst should not create one by allowing an otherwise attractive total to override the failed condition. The method must follow the applicable decision rules.

The final recommendation can be conditional without being evasive. It can identify the preferred admissible option while stating that commitment depends on a particular unresolved gate, or recommend a study because the eligibility evidence is incomplete. The condition should be specific enough to change the action when it fails. A vague promise to manage risks later provides little decision control. A named approval, accountable assessment or demonstrable capability makes the next step reviewable and preserves the boundary between preference and permission.

Keep permission separate from preference

Define the requested commitment and its eligibility conditions before scoring. Assign an evidence owner and disposition to each condition. Run the raw dominance challenge on the admissible comparison set, and explain any retained dominated option. Apply the authorised preference model only after those distinctions are visible. When a gate remains unresolved, recommend a bounded investigation or conditional action that addresses the missing evidence. Do not use a compensatory score to manufacture compliance, operating permission or an exception that the committee lacks authority to grant.

The case for a single integrated score

An integrated score can be easier to communicate than a sequence of checks and may help a preliminary team recognise that all concerns matter. That convenience becomes misleading if a required condition can be traded away by unrelated benefits. The repair is to keep the score as a preference summary while showing eligibility status beside it. A genuinely discretionary concern can remain compensatory. A binding condition requires its own disposition. The extra structure is justified when it prevents a recommendation from confusing attractiveness with permission to proceed.

The app does not certify eligibility

Portfolio Compass checks numerical structure, computes a weighted ranking and reports raw Pareto membership. It does not assess legal compliance, environmental and social performance, safety assurance or delivery authorisation. Those requirements must be resolved through their applicable processes. The ignored-dominance fixture is a logical demonstration with illustrative inputs, not evidence that a particular technology or organisation is inferior. Raw dominance is only as comprehensive as the represented attributes and as credible as their evidence.

Decision implications

- Eligibility concerns the present commitment and requires an accountable evidence owner.
- Raw dominance and a weighted tie can coexist when differences receive no weight.
- A preference score cannot manufacture permission or certify compliance.

MECHANISMS AND WORKED STUDIES

08 The baseline ranking is a conditional result

A frozen illustrative ranking and interpretation table

Trace a conditional recommendation from raw values to contributions

The baseline is a declared reference calculation. It brings together a frozen candidate set, the current evidence interpretations, the stated scales and the authorised weight vector. Its purpose is to make a recommendation inspectable and to provide a consistent starting point for challenges. Calling it the baseline should not imply that it is a neutral view of the world. It contains choices about scope, measurement and preference. Those choices should be visible before the committee interprets the ordering.

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

The frozen illustrative baseline

These are the complete raw values used in the baseline comparison. A lower cost or risk rating is favourable; higher impact or readiness is favourable. The preference score follows the frozen supplied vector and the current-set linear scales.

The frozen illustrative baseline						
Alternative	Cost	Impact	Readiness	Risk	Score	
Reliability architecture	24.00	78.00	85.00	25.00	72.13	
Commercial learning	18.00	68.00	90.00	30.00	62.86	
Energy productivity	42.00	82.00	72.00	32.00	59.32	
Capacity extension	90.00	96.00	52.00	60.00	35.00	

Canonical balanced-baseline input and ranking. Capital cost is illustrative USDm; impact, readiness and risk are supplied ratings. The score has preference-point units.

A baseline presentation should show the raw performance beside the score. Otherwise a committee sees only the aggregation and cannot tell whether an option leads because it is less costly, more strategically attractive, easier to deliver or less exposed to execution difficulties. Raw values also preserve the distinction between a favourable position within the comparison set and an acceptable absolute performance. The least costly candidate in an unaffordable set is still unaffordable. The most ready candidate in an immature set may still need development.

The illustrative baseline uses fictional proposal descriptions and supplied attribute values. The names organise different intervention stories; they do not represent measured projects or verified engineering alternatives. Cost is labelled in monetary units by the app, while impact, readiness and risk are supplied ratings requiring a meaningful rubric in a real decision. This paper uses the fixture to expose calculation behaviour. It makes no claim that a particular intervention has the cost, impact or implementation characteristics assigned to its illustrative name.

Evidence and scope

Multi-criteria analysis can compare options across distinct evaluation criteria.

Evidence: Better Regulation Toolbox: Methodologies for analysing impacts

The next reading is contribution. A contribution is the portion of the total preference score associated with an attribute after transformation and weighting. It is not the monetary benefit attributable to that attribute, nor a causal estimate of how the intervention creates value. Its useful role is explanatory: it shows which represented differences and tradeoffs support the ordering. The committee can then challenge the criterion that actually matters to the result instead of debating a total whose internal structure remains hidden.

A low contribution may have different explanations. The candidate may perform poorly within the current range, the criterion may carry little weight or every candidate may perform similarly. The raw values, normalised values and weights distinguish these explanations. A contribution should not be described as evidence that the underlying outcome is unimportant. It expresses how much that criterion adds to the particular weighted comparison. Its meaning remains conditional on the scale and the other alternatives defining the range.

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

Where the illustrative preference points come from

The ledger decomposes each total into the represented criterion contributions. It permits a reader to trace the leader to the declared preference model rather than treating the total as an unexplained verdict.

Where the illustrative preference points come from					
Alternative	Cost contribution	Impact contribution	Readiness contribution	Risk contribution	Total
Reliability architecture	22.92	12.50	21.71	15.00	72.13
Commercial learning	25.00	0.00	25.00	12.86	62.86
Energy productivity	16.67	17.50	13.16	12.00	59.32
Capacity extension	0.00	35.00	0.00	0.00	35.00

Canonical balanced-baseline contribution ledger, in preference points. Display rounding can make a printed row sum differ slightly from its printed total; the engine retains full numerical values.

Read the leading case conditionally

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

Under these illustrative inputs, Reliability architecture leads at 72.13 preference points, followed by Commercial learning at 62.86. The leader receives 22.92 points from cost, 12.50 from impact, 21.71 from readiness

and 15.00 from risk. Its advantage is distributed across represented attributes. This is an explanation of the supplied preference calculation, not a forecast of the intervention's return or a recommendation to transact.

The capacity proposal illustrates a different profile. It carries the strongest represented impact but weak positions on the other criteria under the current comparison. Its lower baseline position therefore does not mean the impact claim was ignored. The claim contributes according to the declared tradeoff, while the other represented concerns reduce the total relative to competing options. If the organisation believes strategic capacity should dominate this decision, it should examine an explicitly different mandate. It should not pretend that the balanced calculation already embodies that preference.

The commercial-learning proposal illustrates the opposite tension. Lower initial cost and greater readiness can support a strong preference score even when its represented impact is less ambitious. The committee should ask whether the readiness rating reflects the same commitment stage as the larger intervention and whether the impact criterion fairly describes the different scope. A pilot and a capacity commitment may belong in different decision envelopes. The arithmetic can compare their supplied ratings, but the charter must explain why that comparison is legitimate.

The energy proposal may remain relevant despite not leading. A ranked list does not establish that only its first entry has value. If the actual decision permits multiple compatible interventions, different timings or separate funding envelopes, the committee needs a portfolio-selection analysis beyond this app. The baseline can help explain individual preferences. It cannot decide whether an attractive lower-ranked option complements the leader, competes for the same delivery resources or provides a hedge against a shared exposure. Those relationships require an explicit extension of the decision architecture.

Pareto membership adds another reading. An option can remain non-dominated because it possesses some represented advantage, even if its weighted score is lower. That is a reason to examine the tradeoff, not a promise that every non-dominated option should be funded. The set may contain too many legitimate compromises for the available budget. Conversely, eliminating a dominated option does not resolve the choice among the remaining alternatives. Dominance narrows a particular kind of redundancy; weights or another authorised choice rule still express preference among tradeoffs.

The gap between displayed scores should be handled carefully. It is meaningful within the supplied numerical transformation, but it does not have an automatic interpretation as a proportionate difference in strategic value. A score cannot be converted into a probability of success or an expected financial return. The ratings themselves may be uncertain or ordinal in a real workflow. The team should use the gap to identify which changes could matter to the preference calculation, then test those changes against credible evidence and authorised judgement.

The baseline should lead to a proposition that can fail. For example, a fictional committee could prefer the reliability intervention for a bounded next-stage investigation if its operating accountability is confirmed and the balanced mandate remains authorised. That statement identifies the action, its basis and a condition capable of changing it. A proclamation that the top-ranked proposal is best gives little guidance when new evidence arrives. A conditional recommendation tells the next analyst what to revisit and tells the sponsor what the present approval actually permits.

Before presenting the proposition, an analyst should perform a result-to-input walkthrough. Select the leader, identify its important contributions, reopen the underlying raw values and check their evidence perimeter. Then compare the nearest relevant challenger and identify the differences supporting the ordering. This walkthrough can reveal an inconsistent rating definition, a scope mismatch or a mistaken direction of preference. It is more informative than merely confirming that the engine reproduced its own totals. Computational reproducibility and interpretation review serve different purposes.

The committee should preserve the baseline even when it rejects the recommendation. It may choose a different mandate, stage an alternative or request a richer analysis. The original result remains a useful record of what the stated model implied before those changes. Keeping it prevents later revisions from rewriting the basis of the earlier discussion. The archive should include the inputs, engine version and decision rationale, with a separate record for the adopted action. A model output and a committee decision are related artefacts, not interchangeable records.

A conditional baseline recommendation

For the fictional balanced decision envelope, prefer the reliability intervention for the next bounded investigation, subject to confirmation that the operating design and evidence perimeter match the scored proposal. Retain the capacity and energy proposals for explicit mandate or compatibility analysis rather than treating their lower baseline positions as universal rejection. Reopen the recommendation if the weight-use statement changes, a material raw rating is unsupported or a funding constraint alters the feasible action. This original recommendation is an educational illustration; it does not authorise an investment.

The case for presenting only the ordering

A concise ordering can focus a busy committee and reduce the burden of inspecting technical detail. That is reasonable as the entry point to a briefing. It becomes weak as the complete evidence record because it hides the basis of the preference and the conditions that could change it. The repair is progressive disclosure: lead with the conditional action, show the raw comparison and contribution ledger, and keep the full fixture available for challenge. A short meeting should still leave a traceable decision.

A reproducible baseline is not validated project evidence

The fixture verifies the calculation under supplied illustrative values. It does not validate the proposal names, rating rubrics, cost estimates or feasibility of the interventions. The score remains a compensatory preference summary over the current candidate ranges. The baseline does not enforce funding limits, estimate financial returns or capture interactions among selected interventions. Any actual commitment requires the missing evidence and decision processes appropriate to its scope.

Decision implications

- The baseline is a declared reference containing scope and preference choices.
- Contribution ledgers explain aggregation without making causal or monetary claims.
- Archive the calculation separately from the committee action it informs.

MECHANISMS AND WORKED STUDIES

09

Sensitivity should seek preference reversals

A declared weight-shift experiment and reversal-investigation record

Seek the conditions that could overturn the recommendation

Sensitivity should begin with a question capable of changing the recommendation. The analyst asks which reasonable alteration to performance, preference or decision scope could make a different action preferable. A chart of changing scores is useful only if its changes correspond to a decision concern. The purpose is not to decorate the baseline with reassurance. It is to expose the boundary of the recommendation, identify influential assumptions and reveal where further judgement or evidence would matter.

The app provides a specific preference challenge. It changes each normalised criterion weight upward and downward by the fixed percentage-point increment, caps the changed weight at the permitted endpoints and rescales the remaining weights proportionally. The raw performances and candidate universe stay fixed within each challenge. This is a transparent set of individual movements around the baseline. It is not a sample from a population of decision makers, and the fraction retaining the leader is not a probability that the recommendation is correct.

The choice of sensitivity movement matters. A large change can reveal whether a recommendation survives a different priority, but it may be too coarse to locate a nearby reversal boundary. A small change may diagnose local fragility while missing coherent joint changes in several weights. The analyst should describe the app routine exactly and supplement it with separately declared mandate scenarios where relevant. The comparison should not pretend that the selected challenges exhaust all legitimate preferences or all ways the underlying performance evidence could change.

A useful challenge separates the object being varied from the reason for varying it. A readiness weight might change because the committee adopts a different commitment stage. A readiness rating might change because an operating plan is verified. These are different experiments. Altering both simultaneously without a clear scenario narrative makes it hard to identify

why the recommendation moves. The team can use a coherent combined scenario when the circumstances require it, but should preserve simpler comparisons that help explain the causal structure of the calculation.

Evidence and scope

Decision alternatives should be examined with explicit uncertainty and deliberation.

Evidence: NASA Risk-Informed Decision Making Handbook

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

An illustrative contest with opposing strengths

Equal supplied weights produce tied leading preference totals for proposals with different profiles. The engine displays tied alternatives in literal text order. It does not establish substantive priority between the tied alternatives. Literal text order compares code units using JavaScript string comparison, independently of browser language; it is a display convention for equal totals. All co-leading alternatives remain substantively tied under the represented preference.

An illustrative contest with opposing strengths						
Alternative	Cost	Impact	Readiness	Risk	Score	
Flexible service	70.00	80.00	75.00	30.00	50.00	
Operating readiness	90.00	65.00	95.00	20.00	50.00	
Balanced capacity	80.00	75.00	80.00	25.00	47.92	

Canonical close-contest input and output. Capital cost is illustrative USDm; impact, readiness and risk are supplied ratings. Scores are preference points under the frozen vector.

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

The selected individual preference challenges

The raw candidate performances remain fixed. Each row reports the criterion being changed, the resulting weight, the preferred display entry and the resulting leading score. The full ordering is retained in the frozen output for replication.

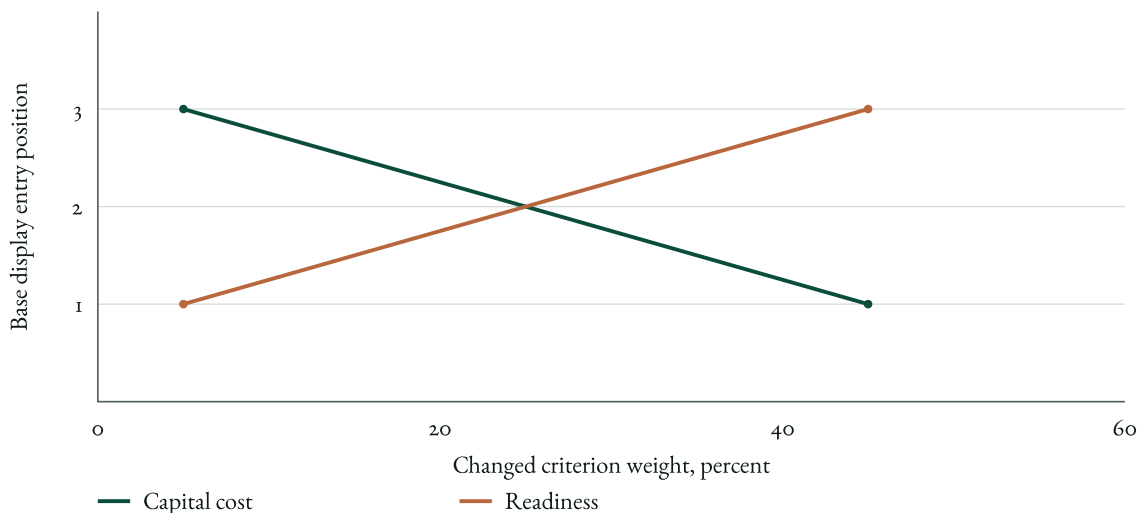
The selected individual preference challenges				
Changed criterion	Shift in percentage points	Resulting weight	Preferred entry	Leading score
Capital cost	-20.00	5.00	Operating readiness	63.33
Capital cost	20.00	45.00	Flexible service	63.33
Strategic impact	-20.00	5.00	Operating readiness	63.33
Strategic impact	20.00	45.00	Flexible service	63.33
Readiness	-20.00	5.00	Flexible service	63.33
Readiness	20.00	45.00	Operating readiness	63.33
Execution risk	-20.00	5.00	Flexible service	63.33
Execution risk	20.00	45.00	Operating readiness	63.33

Canonical close-contest sensitivity output. Weight values are percentages; leading scores are preference points. The challenge set is deterministic and does not represent a probability distribution.

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

Different tradeoffs move the baseline display entry

The illustrative diagnostic selects capital-cost and readiness challenges and shows the position of the original literal text display entry after each movement. The complete table retains every criterion challenge. The base substantive preference is tied; this diagnostic does not turn that display convention into a uniquely preferred option.



Original deterministic drawing from close-contest canonical output pointers. Lower vertical position is a lower displayed rank. Connectors pair discrete tests only; intermediate positions are not computed and the lines do not estimate a continuous rank curve or reversal boundary. Read the preferred names in the accompanying table.

What the experiment actually establishes

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

The illustrative base scores for Flexible service and Operating readiness are both 50.00. The app reports a leader-retention fraction of 50.00 percent across its selected individual challenges. Lower emphasis on capital cost produces Operating readiness as the preferred entry, while greater emphasis on capital cost produces Flexible service. This demonstrates dependence on the declared tradeoff. It does not estimate a chance that either intervention will succeed.

The tie provides a useful adversarial case because a stable literal text display must not be mistaken for a robust substantive choice. The engine needs a deterministic ordering to display and export equal totals. The committee needs a decision rule when the represented preference leaves alternatives indifferent. It might request additional evidence, acknowledge

that either option is acceptable under the declared model or apply an authorised secondary consideration. It should not manufacture precision by presenting the first displayed name as a uniquely justified winner.

The close contest also shows why the identity of the leader is only part of the sensitivity record. Changes may move the original leader below other candidates, change the margin between challengers or affect the ordering of options relevant to a staged programme. The full ranking is therefore retained. A leader-only percentage compresses these distinctions. If a committee plans to keep a reserve option or investigate a runner-up, its sensitivity question should concern that position as well as the first entry.

A more stable reference can still have important limitations. In the balanced fixture, the preferred entry survives the app's selected individual shifts. That is useful evidence about this particular challenge set. It does not mean that a coherent joint change cannot alter the leader, that changed performance inputs cannot matter or that a different candidate universe preserves the scale. The mandate and evidence scenarios in this paper examine some of those other changes explicitly. They should be read alongside the local weight routine, without collapsing them into a single unexplained confidence label.

A recommendation should identify its important reversal conditions in ordinary language. For the close contest, a committee that values lower cost and stronger represented impact differently from operating readiness and risk can favour a different proposal. This is a tradeoff dispute, not necessarily an estimation failure. If both perspectives are legitimate and the commitment is consequential, the next step may be a facilitated mandate discussion. Gathering more engineering detail will not resolve a disagreement about the authorised preference unless the new detail changes the relevant interpretation.

By contrast, a reversal caused by a disputed raw rating can justify targeted evidence work. The analyst should state which proposition needs resolution, what information could change it and what action would follow either conclusion. That structure connects sensitivity to an investigation plan. It also prevents indiscriminate diligence in which every uncertain attribute receives the same attention. The decision-relevant question is whether plausible revision could alter the choice, commitment conditions or scope of the next action.

Sensitivity scenarios need admissibility checks of their own. A weight vector that removes a mandatory concern may be mathematically valid but institutionally illegitimate. A performance scenario that combines favourable assumptions inconsistent with the engineering design may be

impossible. A changed universe containing a proposal outside the decision envelope may answer a different question. The team should distinguish logical model experiments from credible decision scenarios. Both can be educational, but only the latter directly support a commitment discussion under the actual charter.

The team should resist tuning the range of challenges after seeing a preferred outcome. If the report only includes shifts that preserve the sponsor's candidate, the stability display becomes a selection artefact. A predeclared challenge plan can reduce that discretion. It names the kinds of variation, their rationale and the criteria for adding a new challenge when evidence arrives. Unexpected fragility should be retained in the record. A method earns trust by exposing awkward results and explaining their consequence, not by producing reassurance on demand.

Finding no reversal is also a result that needs disciplined interpretation. It may mean the recommendation is stable over a relevant set of changes. It may mean the experiment was too narrow, the tested alternatives are far apart or the changed assumption is not influential. The analyst should say what remained fixed and why the tested movements were relevant. A claim of robustness without that perimeter says little. A bounded statement allows another reviewer to identify a missing challenge and extend the work productively.

The report should keep computation, interpretation and action separate. Computation identifies the ordering under each supplied scenario. Interpretation explains which tradeoff or evidence change produced it. Action decides whether to commit, investigate, redesign or deliberate further. A reversal can lead to different actions depending on whether it arises from legitimate preference disagreement, weak evidence or an invalid option definition. The same movement in a chart does not prescribe the same governance response.

Predeclare the challenge and the consequence

State the baseline recommendation and the uncertainty or preference dispute being tested. Identify what changes and what remains fixed. Check that each decision scenario is credible under the charter, while labelling purely logical experiments separately. Run the canonical comparison, inspect full ordering and ties, then explain the reversal or retention in terms of the changed assumption. Connect that explanation to a specific action: additional evidence, mandate deliberation, redesign or a conditional commitment. Retain inconvenient outcomes and disclose the untested parts of the decision space.

The case for a single stability number

A single leader-retention number can help readers recognise that a recommendation deserves a sensitivity discussion. It becomes misleading when presented as confidence, reliability or likelihood. The repair is to keep the number beside the exact challenge definition and a table showing which scenarios change the preferred entry. For a tied baseline, the display-order convention also needs explicit treatment. The summary remains useful when it directs attention to the detailed explanation rather than replacing it.

Selected shifts do not establish global robustness

The app sensitivity routine is deterministic, local to the supplied vector and limited to individual weight shifts. It does not explore every joint vector, perform probabilistic uncertainty analysis or calculate a reversal boundary analytically. The close-contest fixture is an authored example of conditional preference and tied-score interpretation. It is not a sample of actual investment decisions. A stable leader in the selected tests cannot establish that the full decision is robust to changed eligibility, funding, evidence or interactions.

Decision implications

- A useful sensitivity question can change the proposed action.
- Leader retention across selected shifts is a fraction of deterministic tests, not a probability.
- Tie display order needs a separate substantive decision rule.

MECHANISMS AND WORKED STUDIES

10 Missing values are decisions in disguise

A missing-data policy with consequences

Choose an evidence disposition before completing the calculation

A missing value describes the state of evidence, not the performance of an alternative. An absent cost estimate does not mean that implementation is free. An unresolved risk rating does not mean there is no risk. A blank readiness field does not establish incapacity. The comparison should preserve these distinctions. The canonical app rejects incomplete required numerical inputs and clears the current output until the problem is resolved. That behaviour is a substantive protection against presenting an apparently complete preference calculation on an incomplete evidence basis.

The initial response is diagnostic. Ask whether the value is missing because nobody gathered it, because the concept lacks a usable definition, because the estimate is being withheld or because the decision stage makes a precise value premature. Each cause suggests different work. A missing document may call for a source request. An unclear rubric requires methodological repair. A withheld estimate may require governance intervention. An early-stage option may need a deliberately bounded scenario rather than a false point estimate. A generic command to complete the spreadsheet does not identify these differences.

Missingness also has a perimeter. A sponsor may provide an estimate that omits integration, operating support or a necessary enabling asset. The numerical field is then populated but the decision input remains incomplete. Validation can confirm that a finite value is present; it cannot establish that the value covers the required scope. The evidence acceptance record should therefore check definitions and boundaries as well as numerical availability. An apparently complete dataset can be less informative than a visibly incomplete one if its exclusions are concealed.

The committee should predeclare how incomplete candidates will be handled. Options include deferring the candidate from the present comparison, requesting specified evidence, retaining it as a development option or running explicitly bounded scenarios under assumptions approved for the decision stage. These dispositions are different from assigning a convenient default. The record should explain what is missing, why the disposition is proportionate and what condition would bring the candidate back into the comparison. Otherwise omission can become an unreviewable way to remove an inconvenient proposal.

Evidence and scope

A cost estimate needs a defined baseline and documented assumptions.

Evidence: Cost Estimating and Assessment Guide

A completed missing-evidence disposition register

The fictional review below distinguishes absent values from incomplete scope and ambiguous concepts. It shows decisions the review team can take without inventing numerical performance.

A completed missing-evidence disposition register			
Gap	Diagnosis	Present disposition	Re-entry condition
Capacity proposal cost	Estimate excludes the enabling connection and operating support	Retain as a development option; do not use the partial figure as a complete cost comparison	Sponsor supplies a scoped estimate with exclusions and dependencies identified
Service proposal readiness	Rubric conflates contract signature with operational capability	Pause the rating and resolve the definition for every candidate	Committee approves a common readiness rubric and evidence owners reassess all options
Energy proposal execution risk	Assessment owner has not reviewed integration exposure	Request a targeted assessment before implementation preference is finalised	Risk owner provides the required assessment and its uncertainty perimeter
New commercial proposal impact	Outcome mechanism is stated but the evaluation plan is not defined	Permit a bounded learning design; withhold claims about demonstrated impact	Sponsor specifies the evidence that would support the claimed outcome and the decision following either result

Original completed qualitative example. The app receives no substitute numbers from this worksheet.

A default can be useful for demonstration, but it needs a clear identity. The app's reset values are labelled illustrative so a reader can explore the workflow. They are not estimates for the user's own proposal. A copied demonstration value should not silently become a project assumption. If an analyst elects to use an assumed value in a real early-stage scenario, the record should identify its origin, justification, sensitivity and owner. The decision process, rather than the existence of a populated field, determines whether the assumption is acceptable.

Treating missing performance as the least favourable score is a policy choice with consequences. It may encourage evidence completeness, but it can also penalise innovative or earlier-stage options relative to mature incumbents. That penalty may be appropriate for a near-term implementation decision and inappropriate for allocating investigation resources. The policy should be explicit about the commitment stage and the reason for the penalty. It should not be implemented secretly through a value that looks like measured poor performance.

Treating missing performance as average is another policy choice. It can preserve a candidate in a preliminary screen, but it grants it a position without evidence that the position is warranted. It may also alter the observed ranges and therefore the transformed scores of other candidates. The analyst should avoid presenting such a value as neutral. An assumed central position is still an assumption, with a potential effect on preference. The canonical app does not perform this imputation. Any separate scenario must be labelled and approved on its own terms.

Removing the criterion because one candidate lacks evidence can produce a different distortion. It changes the preference architecture for every candidate and may reward the very option that failed to supply the information. There may be legitimate reasons to simplify a preliminary screen when a criterion cannot yet be assessed consistently, but the committee should acknowledge the resulting blind spot. The original objective or constraint has not disappeared merely because the evidence is inconvenient. A later commitment may still require the missing assessment.

Removing the candidate is equally consequential. The observed-range transformation can change when a proposal leaves the set. The remaining proposals keep their raw performance but may receive different normalised values and scores. The analyst should therefore freeze and document the comparison universe after the missing-evidence disposition. A table should not mix scores computed with an incomplete candidate included and scores computed after it was removed. The reference set is part of the calculation's identity.

Bounded evidence scenarios can preserve an unresolved candidate without inventing certainty. The team states plausible interpretations of the disputed input, gives the basis for each bound and recomputes complete scenarios. This is useful when the bounds are credible and the resulting choice is decision-relevant. It is not permission to choose convenient extremes or call them a confidence interval. The bounds describe an authored uncertainty exercise unless supported by an appropriate estimation process. The evidence-conservative and evidence-optimistic fixtures later in the paper use exactly that limited interpretation.

A bound may remain too broad to support a commitment. If the preferred action changes across credible interpretations and the missing proposition can be investigated, the committee should consider delaying or staging the action. If the missing proposition cannot be resolved in time, the authorised decision maker may need to accept a documented

uncertainty. The model should expose that choice. It should not manufacture a precise recommendation by collapsing an unresolved range into a single unexamined number.

The burden of evidence should be proportionate to consequence. A small reversible learning activity can tolerate a different level of estimation maturity from an irreversible implementation commitment. This does not mean that low-cost activities require no discipline. Their learning purpose still needs a clear question and disposition rule. It means that the evidence request should serve the present decision. Asking for implementation-grade detail before considering an exploratory option can eliminate learning opportunities without improving the immediate choice.

Diagnose a missing readiness rating by trying to complete the activity it is meant to describe. If the sponsor cannot name the operating task, the problem is an undefined scope; another readiness interview will not repair it. If the task is defined but the required capability has never been tested, a demonstration may resolve the proposition. If the capability is demonstrated but unavailable for the proposed period, the missing item is a resource commitment. These routes produce different work: define, test or secure. The evidence request should name that route so a document proving one proposition cannot silently close another.

An unresolved candidate can distort more than its own position. Under observed-range scaling, admitting an assumed extreme value may change the scores of supported candidates by moving an endpoint. Simply hiding the incomplete candidate in the final presentation therefore does not undo its influence. The diagnostic is to inspect whether its supplied placeholder defines any best or worst value, then distinguish two questions: whether the candidate belongs in the present comparison and whether the remaining scale still represents the intended tradeoffs. A development-only disposition should be applied before constructing the implementation comparison, with a fresh interpretation of any changed range.

The practical export check is a round trip between the numerical row and the missingness register. Select the candidate with an assumed readiness value, find the exact proposition that remains unsupported, and verify that the imported row still points to the same scoped activity. A valid input can reproduce an assumption perfectly while an accompanying note refers to a different design stage. Reject that join and repair the correspondence before reuse. Where the numerical schema cannot carry

evidence maturity, the separate register supplies the link; it must distinguish an absent observation from an observed weak performance, since those imply different investigations.

Resolve a gap without disguising it as performance

Identify whether the missing item is an absent value, an incomplete scope or an undefined concept. Match the evidence standard to the present commitment. Assign a disposition and owner before calculating: request evidence, defer the candidate, retain a development option or author a bounded scenario. Recompute a coherent complete comparison set after the disposition. State the unresolved proposition and the action that changes when it is resolved. Keep the numerical fixture linked to the qualitative evidence record so schema validity cannot be mistaken for evidential support.

The case for quick imputation

A preliminary team may need a rough comparison before every estimate is available. Transparent assumed scenarios can serve that purpose when the decision is reversible and the assumptions are deliberately bounded. The problem is silent imputation that presents the substitute as evidence. The repair is to name the assumption, show its consequence under credible alternatives and restrict the conclusion to the preliminary decision. The canonical app requires complete explicit inputs; it does not make an evidence policy on behalf of the user.

Validation cannot establish evidence completeness

Required-field and numerical checks can detect a blank or malformed input. They cannot detect a cost perimeter missing an enabling asset, a rubric applied inconsistently or an estimate unsupported by a source. The app does not impute values, evaluate evidence quality or certify the maturity of a proposal. The fictional disposition register supplies a process design, not findings about actual candidates. A complete numerical scenario remains conditional on its evidence and scope.

Decision implications

- Missing evidence is not favourable or unfavourable performance.
- Disposition depends on commitment stage, scope and accountable ownership.
- Numerical schema validity does not establish evidence support.

MECHANISMS AND WORKED STUDIES

11

Uncertainty is not one confidence multiplier

An uncertainty decomposition worksheet

Keep performance uncertainty distinct from preference and structure

Uncertainty enters a preference model through different channels. The performance estimate may be uncertain, the criterion may be poorly defined, the decision maker may be unsure about a tradeoff or the future context may alter the relevance of an objective. These conditions should not be compressed into a generic confidence rating. They imply different responses. Better measurement can improve a performance estimate; it cannot resolve a legitimate disagreement about priorities. Clearer authority can settle whose preference governs; it cannot repair a missing technical baseline.

Performance uncertainty concerns what the option would deliver or require under the stated scope. A readiness rating may depend on an unresolved operating arrangement. A risk rating may depend on a control whose effectiveness has not been tested. The team should identify the proposition behind the uncertainty and the evidence that could change it. A range of possible ratings is useful only if it reflects plausible interpretations of that proposition. Arbitrary variation creates mathematical movement without a decision meaning.

Preference uncertainty concerns the value attached to performance differences. Participants may understand the evidence yet disagree about whether near-term readiness or strategic impact deserves greater emphasis. The model can show separate vectors, but it cannot estimate the correct collective preference from engineering data. The committee needs an authorised process for adopting a mandate. Where authority remains contested, the recommendation should expose the dispute rather than describe its averaging as a neutral resolution.

Structural uncertainty concerns whether the model contains the right attributes and relationships. A proposal may depend on another investment, combine benefits non-additively or face a binding condition outside the score. Varying its readiness rating does not necessarily represent those relationships. A richer scenario narrative or a different

method may be needed. Sensitivity inside a misspecified model can be perfectly reproducible while leaving the important uncertainty untouched. The analyst should challenge the architecture before interpreting the movement of totals.

Evidence and scope

Risk management requires explicit responsibilities and informed assessment.

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

Match the uncertainty to the response

The categories below are original analytical distinctions for this paper. They organise the work required to improve a decision without implying that every uncertainty can be reduced to a numerical distribution.

Match the uncertainty to the response			
Uncertainty channel	Example proposition	Useful response	Misleading substitute
Performance	An operating plan may not support the supplied readiness interpretation	Verify the proposition and compare credible bounded ratings	Multiply every score by an unexplained confidence factor
Preference	The committee has not authorised the readiness versus impact tradeoff	Facilitate and record separate mandate scenarios	Treat stakeholder disagreement as measurement error
Structure	Benefit requires an enabling intervention outside the option scope	Represent the dependency or revise the decision envelope	Vary a standalone rating while ignoring the enabling commitment
Context	A service obligation may change the relevant decision stage	State the triggering condition and prepare a distinct decision scenario	Present the baseline as universally valid

Original uncertainty-response exhibit. The canonical app calculates supplied scenarios; it does not infer their evidential credibility.

A generic confidence multiplier is tempting because it creates a single adjusted total. It can also conceal several unresolved choices. Does the multiplier represent a chance of delivery, confidence in the estimate, maturity of the evidence or a committee penalty for incomplete work? These meanings are different. Multiplying the preference score by a number does not automatically convert it into expected utility or expected value. Such an interpretation would require a defined decision model and a justified relationship between uncertainty and outcomes.

A confidence penalty can double-count readiness or risk if those criteria already represent the relevant concern. It can also penalise an innovative option for weak evidence even when the present decision is specifically to

investigate it. Conversely, a mature incumbent may receive an unjustified advantage because its documentation is familiar. The analyst should locate the concern in the correct part of the architecture: evidence acceptance, eligibility, a meaningful performance attribute or a separate scenario. The paper does not endorse an automatic confidence-adjusted ranking.

The paired evidence fixtures demonstrate a limited alternative. The raw cost and impact of every candidate remain fixed, as do the supplied weights and candidate names. The logistics proposal receives different readiness and risk interpretations. The change is an authored bounded experiment about disputed evidence, not a confidence interval, a probability distribution or a forecast of actual implementation. The purpose is to ask whether those interpretations change the preference ordering and to identify the result that would matter to the next decision.

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

The supplied evidence interpretations

The same logistics proposal is evaluated under less favourable and more favourable readiness and risk interpretations. Other supplied raw values and the preference vector remain unchanged. The terminology describes the direction of the illustrative assumptions, not estimated likelihoods.

The supplied evidence interpretations			
Attribute of logistics proposal		Less favourable interpretation	More favourable interpretation
Cost		48.00	48.00
Impact		87.00	87.00
Readiness		30.00	85.00
Execution risk		65.00	20.00

Canonical evidence-conservative and evidence-optimistic inputs. Capital cost is illustrative USDm; impact, readiness and risk are supplied ratings. No empirical confidence level is assigned.

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

The leader stays while the runner-up changes

The output should be read by candidate identity, because ranking positions change. The metering proposal remains preferred in both supplied interpretations. The logistics proposal moves ahead of the inspection service in the more favourable interpretation.

The leader stays while the runner-up changes			
Alternative	Less favourable score	More favourable score	Ordering consequence
Metering retrofit	65.83	65.83	Remains preferred
Inspection service	57.25	25.00	Moves behind logistics proposal
Logistics digital twin	35.00	51.67	Moves ahead of inspection service

Canonical paired outputs, in preference points. The comparison changes raw ratings and therefore some current-set normalisation ranges.

Do not promise a reversal the fixture does not show

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

Under both illustrative interpretations, Metering retrofit leads at 65.83 preference points. The inspection score changes from 57.25 to 25.00, even though its raw inputs are unchanged. The logistics score changes from 35.00 to 51.67. The supplied bounds change the runner-up and the scale-based comparison, but they do not overturn the leading choice. This is the stated result of the experiment.

The movement in the inspection score is a consequence of shared range transformation. When the logistics readiness and risk positions improve, the least favourable values defining those ranges change. The inspection proposal occupies a different relative position within the new ranges despite retaining its own raw performance. This explains why a changed score should not automatically be described as a changed estimate of that proposal's quality. The analyst must distinguish performance revision from normalisation movement.

The unchanged leader does not make the evidence request pointless. A committee may care about the runner-up because it wants a reserve intervention, a staged investigation or an alternative if the leader fails an eligibility condition. The logistics evidence could then affect the next action even though it does not affect the preferred entry in the present standalone ranking. Decision relevance depends on the actual action being considered. It should not be inferred solely from whether the first row changes.

The unchanged leader also limits the case for delaying its investigation until all logistics evidence is resolved. If the only present action is a reversible next-stage study of the leader and that action remains appropriate across credible bounds, the committee may have a basis to proceed while investigating the challenger separately. That conclusion remains conditional on eligibility and the relevance of the bounds. It is not a general rule that uncertain competitors can be ignored. The team should state the action and the uncertainty it can tolerate.

The bounded evidence pair can be extended conceptually through an out-of-envelope test without inventing another score. Suppose inspection resolves the installation issue but reveals an integration dependency that neither supplied interpretation included. Selecting the optimistic interpretation would erase the newly discovered mechanism. The correct response is to separate the resolved installation proposition from the unresolved integration proposition and ask whether the existing readiness and risk definitions can represent both without duplication. If they cannot, the uncertainty is structural rather than merely a location within the declared bounds. The paired studies then remain useful historical comparisons but cease to bracket this changed proposal.

Uncertainty can remain irreducible at the commitment date. The committee may have to act without resolving future context or performance fully. A defensible recommendation identifies which uncertainties have been reduced, which remain and who accepts the resulting exposure. It can recommend a staged action, a reserve option or a monitoring trigger. It should not make the score absorb uncertainty that the organisation has never explicitly considered. Clear residual uncertainty is more useful than a precise but ambiguous confidence-adjusted total.

Identify the uncertainty before choosing the experiment

Classify the concern as performance, preference, structure or context. State the proposition, evidence owner and action that depends on it. Use bounded raw-value scenarios for credible performance interpretations, separate authorised vectors for preference disputes and a revised method when the architecture is incomplete. Read the full ordering and distinguish raw revision from scale movement. Report both reversals and non-reversals. Finish with a closure rule for the evidence and an explicit statement of the residual uncertainty the committee is being asked to accept.

The case for a confidence-adjusted total

A single confidence-adjusted score can be convenient when the organisation has a carefully defined policy linking evidence maturity to a preliminary screen. Its legitimacy depends on that policy and on avoiding overlap with represented risk and readiness. It should not be presented as expected utility merely because multiplication is easy. The repair is to disclose the meaning, authority and consequences of the adjustment, and compare it with explicit scenarios. Where those foundations are absent, the unadjusted preference model plus an evidence register is more interpretable.

The paired scenarios do not estimate uncertainty distributions

The less and more favourable fixtures are supplied illustrative interpretations. Their bounds have no empirical coverage level and no probability weights. The app does not sample uncertain inputs, estimate correlations or calculate the value of information. The paired results show conditional rankings only. A stable preferred entry does not settle eligibility, compatibility or funding, and an unchanged raw input can still produce a changed score when the current comparison range moves.

Decision implications

- Different uncertainty channels require different decision responses.
- The paired evidence bounds change the runner-up and retain the same leader.
- A confidence multiplier needs a defined meaning and authority; multiplication alone supplies neither.

MECHANISMS AND WORKED STUDIES

12 Budget selection differs from rank order

A ranking-versus-portfolio counterexample

Define feasible combinations before claiming a funded programme

A preference ranking answers which standalone option leads under the supplied attributes and tradeoffs. A funded portfolio answers which feasible combination of commitments the organisation should undertake within its constraints. Those questions can produce different answers. The highest-ranked option may consume an envelope that would otherwise support a more attractive combination, depend on an unfunded enabling asset or compete with another intervention for scarce delivery capacity. Selecting entries down a ranked list until the money runs out is a policy, not a general solution to portfolio allocation.

The distinction begins with the unit of decision. A standalone proposal can have an interpretable cost, impact, readiness and risk profile. A combination requires information about how these profiles interact. Benefits may overlap, delivery resources may be shared and costs may change when components are procured together. Summing standalone preference scores assumes more than the ranking model establishes. The committee should not treat the score as an additive portfolio benefit merely because it is expressed numerically. Its scale depends on the candidate set and its components mix unlike considerations.

A budget constraint is different from a preference for lower cost. The liquidity-mandate fixture changes the cost weight, which alters the attractiveness of standalone alternatives. It does not introduce an expenditure ceiling or enforce affordability. A very costly option can still receive a strong score if its other represented attributes compensate. Conversely, an affordable option can rank poorly. If the present decision has a binding envelope, feasibility must be checked explicitly in the selection method. Increasing the cost weight cannot substitute for that check.

Indivisibility matters when an intervention cannot be purchased in arbitrary fractions. A committee cannot always take part of the leading proposal and allocate the remaining envelope to another. A minimum viable scope, contractual package or enabling requirement may make the commitment indivisible. Greedy ordering can leave unusable capacity in the envelope or exclude a useful combination. The analyst should record whether each candidate is divisible, stageable or indivisible and whether a reduced version has its own credible performance profile. A fractional arithmetic solution is not automatically an implementable plan.

Evidence and scope

Economic appraisal considers value beyond investor financial profitability.

Evidence: [The Economic Appraisal of Investment Projects at the EIB, Second Edition](#)

Information a funding decision needs beyond the ranking

This original exhibit identifies selection facts that are not fields in Portfolio Compass. They must be established before treating standalone preferences as a funded programme.

Information a funding decision needs beyond the ranking			
Selection feature	Question to resolve	Decision consequence	Ranking limitation
Binding capital envelope	What expenditures and timing belong inside the approved envelope?	Some combinations are infeasible regardless of their preference totals	A cost weight is not a ceiling
Indivisible scope	Can the intervention be reduced without changing its outcome and operating design?	Selection may require whole packages or separately evaluated stages	A ranked entry has no automatic fractional version
Mutual exclusion	Do proposals serve the same purpose in incompatible ways?	Funding both may duplicate or conflict with the service design	Standalone scores do not encode exclusion
Complementarity	Does benefit depend on another selected intervention?	A lower-ranked enabler may be necessary for a desirable package	Standalone performance does not establish package value
Delivery capacity	Which accountable teams and capabilities are required concurrently?	An affordable programme can still be infeasible to execute	Readiness ratings do not allocate shared resources
Recurring commitment	Which operating and maintenance obligations follow implementation?	Initial capital affordability can conceal later financial exposure	The app cost field does not calculate lifecycle cash requirements

Original portfolio-feasibility architecture. No allocation is computed by this paper or the paired app.

Mutually exclusive alternatives require particular care. If several proposals represent different ways to deliver the same service, their standalone ordering can support a choice within that family. It cannot justify selecting all of them. The committee should identify the family and the exclusivity rule before constructing a funded programme. Otherwise a collection of individually attractive options may become an incoherent set. The programme must describe what is implemented, which service it delivers and which alternatives it displaces.

Complementary interventions create the opposite problem. An enabling data architecture may have modest standalone impact but make several subsequent interventions possible. A high-ranked operating intervention may require it. Selecting only the operating intervention would then violate the scope underlying its performance claim. The analyst should examine packages or explicit dependencies, with benefits attributed

consistently. A dependency graph can clarify the structure, but it does not by itself estimate the value of the package. The necessary relationship and the preference for the resulting outcome both need attention.

Shared benefits can make a sum of scores especially misleading. A monitoring proposal and a process redesign may both claim the same reduction in disruption. Funding them together does not necessarily double that outcome. The programme should identify the benefit mechanism and how the interventions interact. The issue is not solved by assigning each a lower subjective impact rating after the fact. A package-level evidence model should explain the incremental outcome and the assumptions supporting it. The standalone ratings remain useful only for their original comparison purpose.

A symbolic counterexample helps clarify the logic without pretending to solve a funded case. Suppose the leading proposal uses most of an envelope and leaves no room for another indivisible intervention. A pair of lower-ranked compatible proposals fits the envelope and creates a more desirable programme outcome under a separately defined package evaluation. A greedy selection from the standalone ordering will choose the leader; a constrained package comparison may choose the pair. Nothing in the standalone ranking proves which programme is preferable, because the required combination values and constraints were never represented.

The analyst should not respond by dividing the preference score by cost and declaring the result a universal efficiency measure. The score includes cost already, mixes transformed attributes and changes with the reference set. Dividing it again can double-count cost preference and create a ratio without a meaningful unit. A benefit-cost or cost-effectiveness analysis requires a defined outcome measure and consistent cost perimeter. The organisation may choose such a method, but it is a different analysis with its own assumptions and evidence requirements.

Time introduces another layer of feasibility. An intervention can fit a total envelope but require expenditures before funding is available, or compete with another proposal during the same delivery window. A programme may be feasible when staged and infeasible when launched concurrently. The selection method should represent the relevant timing and resource obligations explicitly. A readiness rating can summarise some delivery maturity, but it cannot replace a schedule or a cash requirement profile. The committee should know which temporal constraints remain outside the preference model.

A preliminary ranking can still make the allocation process more manageable. It can identify proposals worth detailed feasibility work, reveal tradeoffs among standalone options and organise questions for package design. Its role should be stated as screening or preference explanation rather than optimisation. The team can retain the ranking while building a separate constrained comparison of plausible packages. This preserves the work already done without letting its convenient format dictate the later decision rule.

Before handing individual options to package analysis, construct a compatibility ledger in words. For each proposed pairing, identify the shared resource, the reason simultaneous delivery could fail and whether sequencing would remove the conflict. A common operating team may create a temporary capacity conflict; two designs for the same asset may remain mutually exclusive; an enabling interface may make joint delivery more valuable. These relations cannot be recovered from the four standalone ratings. Transfer consistent raw commitments and their scope, while withholding preference totals as package benefits. This ledger defines the representation the allocation analysis needs without pretending that the present app has optimised it.

Use the next funding round as a diagnostic of the reason an option was left out. A deferred candidate excluded by an annual envelope should return when that envelope changes, subject to current evidence. A candidate excluded because the selected programme now supplies its function may no longer add anything even if money becomes available. A rejected design with an unresolved incompatibility should not re-enter simply because its old score was high. These are queue, redundancy and feasibility branches. The carry-forward register should preserve the applicable branch, since retaining only the ranking would confuse a temporary constraint with the disappearance of the underlying need.

A defensible allocation brief describes the selected programme, its feasibility and the tradeoffs accepted, then explains how the standalone analysis informed it. It does not claim that the ranking selected an optimal funded portfolio when no constraint model was run. If the organisation has not yet established package benefits or shared-resource requirements, the recommendation can be to perform that analysis before commitment. That is a concrete next action, not a failure to produce a convenient answer. The method boundary protects the meaning of the recommendation.

Hand a ranking into a constrained selection process

Define the funded decision unit and envelope, then classify indivisibility, mutual exclusions, dependencies, shared benefits and resource timing. Preserve the standalone preference record as a screening input. Develop credible packages or a constrained allocation method with outcome and cost definitions suitable for combination. Check feasibility independently of preference, compare the feasible actions and record why unfunded candidates are deferred or rejected. If the additional evidence is absent, authorise the targeted selection analysis before claiming an optimal programme.

The case for selecting down the list

A simple ranked-list policy can be practical when proposals are independent, similarly scoped, individually affordable and governed by an explicitly accepted funding rule. It can reduce administrative complexity and make the process predictable. Its justification is those conditions and that policy, not a general optimisation property of the scores. The repair is to test for exclusions, complements and binding resources before applying the rule, then disclose that it is a selection convention. Where interactions or indivisibility matter, a package comparison is warranted.

No funded portfolio is solved here

Portfolio Compass computes standalone preference scores, raw Pareto membership and selected weight challenges. It does not enforce a capital envelope, optimise combinations, allocate delivery resources or value synergies. The symbolic counterexample identifies a logical failure mode without supplying numerical package utilities. The methodological conclusion is that ranking and allocation require different information. An actual funded recommendation needs a credible constrained selection analysis beyond this app.

Decision implications

- A cost preference does not enforce a funding ceiling.
- Standalone preference totals are not automatically additive package benefits.
- Indivisibility, exclusivity, dependencies and timing can defeat a greedy ranking policy.

GOVERNANCE AND APPLICATION

13 Dependencies determine strategic options

A dependency-aware capital pathway

Make enabling relationships and later choices explicit

An intervention can deserve attention because it enables a later choice, even when its immediate standalone benefit is modest. A data foundation, operating capability or connection asset may expand the feasible actions available to the organisation. The committee should identify that enabling mechanism explicitly. Calling a proposal strategic does not establish the mechanism, and assigning it a generous impact rating does not resolve the dependency. The decision architecture should show what becomes possible, under what conditions and with what further commitments.

A dependency is a required relationship, not merely an attractive association. The team should ask whether the later intervention can deliver its stated outcome without the enabler, whether an alternative enabler exists and whether the dependency concerns technical function, permission, information or organisational capability. These distinctions matter. A technical prerequisite may require a specific sequence. An information prerequisite may permit a study before implementation. An organisational prerequisite may need an accountable capability-building plan rather than an asset purchase.

The scope of the candidate should include dependencies consistently. If the capacity proposal's readiness assumes an enabling connection that is not funded or delivered, its standalone rating may be misleading. The analyst can redefine the candidate as a complete package, make the dependency an explicit condition or compare staged actions. The choice depends on the present commitment. What should not happen is that the proposal receives the benefit of the enabler in its ratings while omitting the enabler's cost and delivery burden from the decision record.

Dependencies also affect the interpretation of timing. A technically mature component may be unable to proceed because an upstream permission or capability remains unresolved. Conversely, a preliminary study may proceed in parallel with enabling work if its outputs do not

depend on completion. The dependency record should identify which activity requires which predecessor and why. A single readiness rating cannot carry all of this sequence information. It can summarise a judgement for a standalone preference screen, but a commitment plan needs a more explicit architecture.

Evidence and scope

A credible schedule needs logically linked activities and dependencies.

Evidence: Schedule Assessment Guide: Best Practices for Project Schedules

A completed enabling-dependency record

The fictional record distinguishes required capability from the next commitment. It is a qualitative design illustration, with no project schedule or option value calculated.

A completed enabling-dependency record			
Proposed action	Required condition	Dependency type	Next reviewable commitment
Operational analytics deployment	A governed data feed and an accountable response team	Information and organisational capability	Verify data access and operating responsibility before implementation approval
Capacity intervention	A compatible connection design and required approvals	Technical and permission	Authorise a scoped interface study that identifies the enabling commitment
Inspection service expansion	A demonstrated workflow for acting on findings	Operating capability	Test the response process before scaling coverage
Energy intervention	An integration design consistent with the operating system	Technical	Complete the integration assessment and revise the candidate scope if needed
Commercial learning activity	A question with a decision that changes after learning	Information	Approve the bounded learning protocol and its disposition rules

Original completed dependency worksheet. Each row identifies a relationship that a standalone preference score cannot enforce.

Strategic-option language should identify a choice that is preserved or created. An early investigation may allow the organisation to decide later whether to commit, redesign or stop. An enabling asset may permit alternative downstream uses. A modular contract may reduce the cost of changing direction. These features can matter, but they need a concrete

description. A proposal does not gain defensible option value merely because its future is uncertain or its sponsor calls it flexible. The organisation should know what right it obtains and what obligations accompany it.

The distinction between an option and an obligation is especially useful. An enabling investment may make a future action possible while also creating operating commitments or political expectations that make stopping difficult. A pilot may be technically reversible but institutionally become a promise to scale. The committee should examine both the formal choice and the practical pressure to continue. A staged plan that cannot credibly stop after an adverse result does not provide the same flexibility as one with an enforceable exit and a clear decision rule.

Information can be an enabling benefit when it changes a later decision. A study should specify the uncertainty it addresses, the possible findings and the actions that follow those findings. If the organisation plans to implement regardless of the result, the study may serve design or assurance rather than choice. Those purposes can still be valuable, but they should be named correctly. The claim that an investigation creates an option needs a subsequent decision that remains genuinely open.

A readiness penalty can obscure such enabling value. An exploratory intervention may be deliberately immature because its purpose is to learn before a larger commitment. Comparing it directly with a mature implementation proposal under a readiness-heavy vector can answer the wrong question. The committee should determine whether it is choosing an implementation now, allocating development effort or selecting a sequence of commitments. Different envelopes can make the staged option interpretable without inflating its standalone ratings to compete with a proposal at another stage.

Delayed benefit should also be explicit. An enabler may consume resources now and contribute only when downstream work succeeds. The preference architecture should not credit the full eventual programme benefit as though it were an immediate standalone outcome. A package appraisal or staged decision model can represent the relationship more faithfully. The paired app does not discount future outcomes or calculate real-option values. This paper treats enabling logic as a qualitative decision-design concern and states when another analytical method is needed.

Dependencies can create concentration risk. Several desirable interventions may rely on the same scarce capability, permission or supplier interface. The programme may appear diversified by proposal

name while sharing a single point of failure. A dependency record should therefore identify common enablers as well as pairwise relationships. This helps the committee distinguish a collection of activities from a resilient programme. It does not automatically imply that redundancy is desirable; the cost and feasibility of alternatives still require assessment.

An enabler can also become a bottleneck in the approval process. If every downstream proposal assumes it will be available, the organisation may approve commitments before resolving who funds and owns it. The record should assign the enabling action an accountable sponsor and explicit status. Downstream recommendations can then state whether they are conditional on it, include it in scope or remain valid without it. This removes an ambiguity that a favourable preference score cannot detect.

Mutual dependence requires a different response from a simple sequence. If proposals only create benefit together, evaluating each as a self-sufficient alternative may understate or misstate the programme. The analyst should consider a package with its own outcome evidence, costs and delivery conditions. If the relationship is uncertain, targeted investigation may be the appropriate next commitment. The committee should not solve the problem by giving each proposal the full package impact. That double credit can make aggregation attractive while obscuring the actual unit of value.

Test a claimed dependency by holding its required function fixed and changing the implementation. If the downstream proposal needs timely operating data, ask whether that need requires the proposed platform or can be met by a narrower interface or existing capability. A substitute that meets the function disproves indispensability of the first design, though it may introduce different support or integration costs. If every credible substitute fails the same service condition, the dependency has stronger support. The resulting record distinguishes a required function, an incumbent design and a preferred delivery route. That distinction can preserve strategic flexibility before detailed commitments make substitution expensive.

The committee should decide how to stage the commitments in light of these relationships. It may approve an enabling study, reserve a downstream option conditionally or request a package comparison. Each action should identify what is authorised now and what remains undecided. The staged record prevents later teams from treating a preliminary preference as approval of the entire programme. It also gives sponsors a concrete path to the next decision, with evidence and conditions rather than an open-ended promise of support.

Preserving options has a cost. Additional studies, modular arrangements and reserve capabilities can consume capital, staff time and management attention. A proposal to remain flexible should explain why the expected decision opportunity matters and what burden is accepted to preserve it. The qualitative option architecture does not establish that every staged design is preferable. It identifies the question that a fuller appraisal must answer. Flexibility should be treated as a feature with conditions and costs, not as a universally favourable label.

Describe the enabling choice before assigning strategic value

Identify the downstream action that becomes possible and the condition that enables it. Distinguish technical, permission, information and organisational dependencies. Check whether the candidate scope includes the required cost and delivery burden. Describe what choice remains open after the present commitment, what would permit stopping and what obligations make exit difficult. Where benefit is joint, compare a package or a staged decision rather than inflating standalone scores. Authorise the next bounded commitment with an owner and a condition for moving to the following stage.

The case for a broad strategic-impact rating

A broad impact rating can capture enabling value in an early screen when a detailed package model would be premature. It is useful if the rubric requires a concrete mechanism and distinguishes immediate from contingent benefit. Its weakness is that a narrative about future possibilities can absorb almost any attractive claim. The repair is a short dependency and option record beside the rating. If the claimed impact depends on material commitments outside scope, the committee should treat the standalone score as provisional and commission the appropriate package analysis.

No real-option valuation or dependency enforcement

The app does not model a dependency graph, contingent exercise decisions, staged cashflows or real-option values. The qualitative worksheet describes mechanisms an actual decision process must verify. It does not establish engineering necessity or economic value. A high preference score cannot confirm that an enabler will be delivered, that a later choice remains open or that an exit is feasible. Those claims require evidence and, where consequential, a method capable of representing the staged programme.

Decision implications

- An enabling claim needs a concrete downstream action and accountable condition.
- A stage creates flexibility only if the later choice remains genuinely open.
- Joint and delayed benefits require a package or staged analysis beyond standalone ranking.

GOVERNANCE AND APPLICATION

14 Different audiences can prefer different portfolios

A stakeholder scenario comparison

Preserve disagreement until authority and tradeoffs are explicit

Different stakeholders can understand the same evidence and legitimately prefer different actions. A strategy sponsor may prioritise capacity outcomes, an operating owner may prioritise executable improvement and a finance function may emphasise the burden of initial capital. The analyst should distinguish this preference disagreement from disagreement about performance. The model can calculate the consequence of each supplied vector. It cannot determine which participant has authority to govern the commitment or whether an apparent compromise is legitimate.

A stakeholder scenario needs a mandate, not merely a role label. It should state the objective, commitment stage, represented interests and decision rights associated with the perspective. A finance-labelled vector does not automatically represent the organisation's financial policy; a delivery-labelled vector does not establish what the delivery team actually believes. Actual participants must supply and authorise their interpretation. The illustrative mandates in this paper are authored alternatives designed to test conditional preference, rather than survey findings or descriptions of an institution.

Authority matters because averaging can conceal a decision that still needs to be made. If a board has delegated implementation priorities to an operating committee within a strategic envelope, a simple average of unrelated stakeholder vectors may override that delegation. If affected groups have a consultation or consent role, an average may dilute rather than satisfy the relevant process. The governance review should identify whose preferences inform the analysis, whose approval is required and who can adopt the final mandate. The calculation follows that structure; it does not create it.

The team should also distinguish advisory diversity from collective choice. Separate stakeholder scenarios can reveal which proposals remain attractive across perspectives and which depend on contested priorities.

That is useful diagnostic information even if the committee ultimately adopts a single authorised vector. A collective choice process may then negotiate, deliberate or use a formal rule. Each approach has its own legitimacy conditions. Presenting the arithmetic mean of weights as consensus skips the work of establishing whether participants accept the tradeoffs embedded in that mean.

Evidence and scope

Boards address strategy, major expenditures, risk and access to information.

Evidence: G20/OECD Principles of Corporate Governance: Responsibilities of the Board

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

Different illustrative mandates select different leading proposals

Every raw candidate value and the candidate universe remain identical. Only the supplied weight vector changes. Each leading score has meaning within its own declared mandate; it is not evidence that one mandate creates more organisational value than another.

Different illustrative mandates select different leading proposals				
Authored mandate	Preferred entry	Leading score	Leader retention in selected shifts	Decision reading
Balanced standalone preference	Reliability architecture	72.13	100.00	Several represented concerns support the leader
Strategic-impact emphasis	Capacity extension	75.00	75.00	The impact tradeoff changes the preferred entry
Initial-capital preference	Commercial learning	89.29	62.50	Lower initial cost receives greater preference; no budget ceiling is enforced

Canonical balanced-baseline, impact-mandate and liquidity-mandate outputs. Scores are preference points; retention fractions concern the app selected individual weight shifts.

A reversal in mandate is not a revision in performance

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

The illustrative balanced mandate prefers Reliability architecture, the impact-led mandate prefers Capacity extension and the initial-capital mandate prefers Commercial learning. No candidate performance changes across those calculations. The comparison establishes that different

supplied tradeoffs can produce different leaders. It does not establish which tradeoff a real committee should adopt, and the initial-capital case still does not solve a constrained funding allocation.

The mandate reversal is an opportunity for a focused conversation. Participants can inspect the raw performance and explain which swing they value differently. If a strategy sponsor prefers the capacity proposal because the decision is intended to secure a future service capability, the committee can examine whether the readiness-heavy baseline fits that purpose. If an operating owner prefers the reliability proposal because the present commitment concerns a near-term service obligation, the team can test whether the alternatives belong in different stages. The conflict may require a clearer charter rather than a compromise weight.

The finance perspective needs particular care. Greater cost preference can reflect a desire to preserve capital, but it cannot enforce an approved envelope or account for expenditure timing by itself. A stakeholder may believe that a proposal is unaffordable, which is a feasibility claim, rather than simply less attractive on cost. The facilitator should place that claim in the appropriate part of the decision process. Mixing an affordability constraint into a cost weight can obscure the reason an option is rejected and leave the supposed constraint open to compensation.

A common performance table helps keep the discussion disciplined. Participants should be able to disagree about priorities while using the same scope and evidence interpretations. If they also dispute raw ratings, the team should record that as a separate channel of disagreement. A combined stakeholder scenario can be legitimate, but it should identify both changes explicitly. Otherwise a preference dispute may be misread as a technical difference, and participants may talk past one another while comparing totals generated from different premises.

The facilitator should look for agreements that do not require identical weights. Participants may agree on eligibility, acknowledge that an option is dominated on represented attributes or support a bounded investigation before implementation. They may also agree on the evidence needed to resolve a performance dispute. These agreements can support progress while priorities remain contested. The aim is not to force every participant into a single vector. It is to identify what action is authorised now and what disagreement still governs a later commitment.

A robust shared preference can be useful when it is actually demonstrated over relevant perspectives. If the same option leads under several authorised scenarios, the committee has evidence that the present disagreement does not alter that particular preference. It should still state

the perspectives tested and the remaining constraints. In the frozen mandate experiment, the leaders differ, so the paper cannot claim cross-mandate agreement. The result directs attention to the decision envelope and authority that select the applicable vector.

Compromise can still be appropriate. Participants may deliberately adopt an intermediate vector after understanding the tradeoffs and accepting the resulting action. That is a negotiated judgement, not an objective discovery by averaging. The record should explain why the compromise is acceptable, what interests it represents and what conditions would reopen it. The analyst can compute its consequence once it is supplied. The committee should avoid retrospective claims that the compromise was inevitable or that dissent vanished because a number was agreed.

Dissent should be preserved in a form that can improve the decision later. A dissenting participant can state the alternative mandate, the preferred action under it and the trigger that would make its concern decisive. For example, a capacity-oriented perspective may become governing if the organisation formally changes the commitment stage to long-term service expansion. A delivery-oriented concern may become decisive if a required capability fails. This is more useful than recording only that someone opposed the final choice. It gives future reviewers a proposition to examine.

Power asymmetry can shape elicitation. Participants may hesitate to challenge a senior sponsor's weights, or a facilitator may assume that agreement in the room means informed acceptance. A credible process creates a way to state alternative scenarios and unresolved concerns without treating them as obstruction. It can ask participants to explain their own rationale before seeing a consolidated result. The exact process depends on the institution, but the methodological requirement is clear: the model should not make contested authority look like technical consensus.

A mandate dispute can be diagnosed by asking each participant to interpret the same opposed pair rather than another complete ranking. The impact-led participant explains the advantage for which greater commitment is acceptable; the liquidity-led participant explains which commitment difference it refuses to exchange. If both describe the same decision horizon and resource, the disagreement concerns preference. If one is selecting a development study while the other is protecting an implementation envelope, the apparent weight dispute concerns different actions. Repair that mismatch first. The supplied mandate reversals reveal conditional preferences; they cannot establish that averaging the vectors would create a coherent shared action.

A retained minority scenario is most useful when attached to a prospective switch condition. An impact-led alternative might become relevant when the institution moves from near-term delivery to an authorised expansion decision. That is different from a minority claim that the current impact evidence is understated: the latter requires a corrected performance interpretation, not merely another vector. Preserve the alternatives with these different triggers. At review, test whether the triggering change actually occurred before presenting the old minority ordering as current advice. This keeps dissent actionable and prevents an archived scenario from becoming a convenient substitute for resolving the disputed proposition.

A completed disagreement and authority record

This fictional record shows how a committee can preserve distinct viewpoints while authorising a specific next action. It is an original governance design, not a report of an actual meeting.

A completed disagreement and authority record			
Issue	Distinct interpretation	Authority or evidence question	Recorded disposition
Present decision stage	Operating owner seeks a near-term improvement; strategy sponsor seeks future capacity	Does the approved envelope concern implementation now or development of a later programme?	Use the present charter for the bounded investigation and retain a separate future-capacity scenario
Initial capital burden	Finance participant seeks capital preservation	Is the concern a preference or a binding envelope?	Check affordability separately before adopting the preference vector
Collective weights	Participants understand the same raw table but value swings differently	Who may authorise the governing tradeoff?	Committee adopts a stated vector with recorded rationale rather than calling an unexplained average consensus
Dissent	Capacity sponsor prefers the impact-led scenario	What context would make that scenario governing?	Reopen the mandate if the formal decision envelope becomes strategic expansion

Original completed facilitation record. The illustrative mandates are model experiments; the record supplies no claimed stakeholder measurements.

Record whose tradeoff governs and why

Separate disagreement about performance from disagreement about preference. Give each scenario a mandate, decision stage and authorised perspective. Use a common evidence perimeter where possible, calculate the full scenarios and inspect whether the action changes. Identify decision authority before consolidating vectors. Preserve substantive dissent with its preferred action and triggering condition. Adopt the governing vector through the applicable institutional process, and state why it applies to the present commitment. Arithmetic can reveal disagreement; governance must resolve the authority to act.

The case for averaging stakeholder weights

An average can be a transparent collective rule when participants understand and authorise it, the institution accepts equal or specified representation and the resulting tradeoff remains meaningful. Its weakness is the frequent assumption that averaging automatically creates consensus or legitimacy. The repair is to state the aggregation rule before viewing the preferred outcome, show separate scenarios and record who accepts the consolidated mandate. If authority or non-compensable interests are unresolved, averaging should not be used to conceal the remaining decision.

Authored mandates are not stakeholder evidence

The frozen vectors are illustrative supplied preferences. They are not survey data, votes, negotiated agreements or empirical descriptions of finance, strategy or operating functions. The app calculates each vector independently and does not aggregate people, assess representation or establish authority. A change of leader proves conditional dependence within the model. It does not determine which mandate is legitimate or which institutional process must govern a real commitment.

Decision implications

- A mandate reversal changes preference without changing raw performance.
- Averaging weights is a collective rule requiring authority, not automatic consensus.
- Preserve dissent as a scenario with a rationale and triggering condition.

GOVERNANCE AND APPLICATION

15

A score can be gamed before the meeting

An adversarial scoring audit

Protect the objective from the incentives created by its metrics

A preference process creates incentives around what it measures. Sponsors who understand the criteria may present their proposals in terms that improve the ratings, emphasise convenient evidence or define scope to avoid an unfavourable comparison. This is a plausible design concern, not an empirical claim about how often organisations behave. The review process should anticipate it because the calculation cannot distinguish a genuinely improved proposal from a strategically revised description. Evidence acceptance, scope control and independent challenge therefore belong beside the score.

The most direct failure is unsupported rating improvement. A sponsor may assert that a delivery plan is ready without demonstrating operating responsibility or may describe an uncertain outcome as high impact. The analyst should ask which evidence changed and whether the same rubric was applied to every candidate. A revised number should travel with a revised proposition and source. If the underlying evidence is unchanged, the record should explain why the interpretation changed. Otherwise the preference result can improve while the actual decision basis does not.

Scope splitting can create a subtler problem. A large intervention may be divided into a modest apparently ready component and an omitted enabling commitment. The scored component receives a low cost and favourable readiness while retaining a narrative about the whole programme's benefit. The repair is a consistent perimeter. The candidate should include the dependencies required for the claimed outcome or explicitly represent the present commitment as a stage with a narrower benefit. A small-looking option is not automatically comparable with a complete implementation package.

Metric substitution occurs when participants optimise the rating rather than the objective. A readiness rubric based only on document completion may reward paperwork without operating capability. An impact rubric based on activity volume may reward more activity without

the desired outcome. The committee should inspect the mechanism linking the proxy to the objective and the conditions under which that link fails. The problem cannot be corrected solely by assigning a different weight. The underlying criterion definition may need revision.

Evidence and scope

Rated criteria distinguish price, lifecycle cost and non-price attributes.

Evidence: Project Procurement: Rated Criteria

A completed manipulation challenge register

The fictional challenges identify evidence and architecture checks rather than allegations about an actual sponsor. Each row states a repair that preserves the decision meaning of the inputs.

A completed manipulation challenge register			
Potential manoeuvre	What becomes misleading	Challenge question	Review disposition
Split out an enabler	Cost covers a component while impact describes the programme	Can the stated outcome be achieved within this candidate scope?	Revise scope or narrow the claimed outcome before scoring
Declare readiness from documents	Administrative completion substitutes for executable capability	Which operating capability is demonstrated by the evidence?	Apply the common capability rubric and identify unresolved owners
Add a poor reference proposal	Range endpoints move and inflate relative scores	Was this candidate admitted under the predeclared universe rule?	Review inclusion and recompute a coherent frozen set
Change weights after seeing the leader	Outcome tuning replaces authorised tradeoff elicitation	What changed in the mandate or swing interpretation?	Retain the original vector and record the justified revision separately
Count the same benefit under several criteria	A single mechanism receives several preference contributions	Which distinct outcomes do the criteria represent?	Rework the objective map or disclose intentional multiple weighting
Remove an inconvenient criterion	Missing evidence becomes a blind spot for every candidate	Why has the underlying concern ceased to matter?	Resolve the evidence policy rather than silently simplifying the model

Original adversarial review exhibit. It demonstrates how a transparent model can still require governance around its supplied values.

Observed-range normalisation makes candidate admission especially consequential. Adding a reference proposal can change the scale even if it never becomes a serious choice. The range-artifact fixture intentionally adds a proposal that is worse on every represented raw attribute than

relevant comparators. Original candidate values and weights remain fixed. The exercise demonstrates how the scores of retained candidates can move when the comparison universe changes. It does not claim that the leader must reverse or that any actual sponsor has used this manoeuvre.

A scale movement that does not change the leader

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

In the illustrative baseline, Reliability architecture leads at 72.13 preference points. After the additional reference proposal changes the ranges, the same candidate leads at 89.98. The new reference proposal has a score of 0.00 and is dominated on the represented raw comparison. The original leader remains the leader. The higher score is a consequence of the new comparison scale, not evidence that its raw performance improved.

An inclusion rule limits this discretion. The committee should define what qualifies a candidate for the present universe before reviewing its effect on the score. A reference proposal can be legitimate when it is a credible available option within the envelope. It should not be admitted solely to improve the relative position of a preferred candidate. The review should document the source and rationale for the option, its scope and its disposition. If it is removed, recompute the entire comparison rather than retaining scores from the altered scale.

Independent rating review can help without pretending that judgement disappears. A second reviewer should examine a selected rating's evidence, rubric and perimeter before seeing the preferred outcome where practical. The review asks whether the interpretation is supportable and consistently applied. It does not need to invent an ostensibly objective number. Differences between reviewers can reveal ambiguity in the rubric or a genuinely uncertain proposition. Those differences should feed an evidence scenario or a definition repair rather than an unexplained average.

Version control is another useful restraint. A revised input should be stored as a new scenario with its reason and authorisation, while the original remains available. This makes outcome-driven changes easier to identify and legitimate corrections easier to defend. The record should connect the numerical difference to the changed evidence or mandate. A sequence of scenario files without reasons can still obscure the decision history. The platform's input-only vault supports reproducibility, but the committee record supplies the justification for each revision.

The team should test for favourable omissions as well as unfavourable values. A proposal may omit a necessary operating obligation or describe risk only within a narrow component. Required numerical fields cannot

detect that omission. A perimeter checklist should ask which commitments, interfaces and dependencies are included and which are excluded. The same checklist should apply to every candidate. Consistency matters because a complete proposal may appear more costly or less ready than an incomplete one precisely because it acknowledges what implementation requires.

The analyst should also inspect how the story changes between proposal and briefing. A cautious estimate in a source document can become a confident rating in the model and a definitive statement in the recommendation. Each translation can lose qualifications. The evidence record should retain the original proposition, the interpretation supporting the rating and the scope of the resulting recommendation. A reviewer can then identify whether the briefing has overstated what the evidence supports. A polished output should not erase the uncertainty that matters to the commitment.

Governance can overcorrect. A rigid anti-gaming process may reject legitimate redesign or improved evidence because it treats every revision as suspicious. The purpose is to distinguish justified change from unexplained outcome tuning. A sponsor should be able to improve a proposal, narrow its scope or supply better evidence. The process should reward that work by evaluating the revised candidate transparently. It should preserve the previous version so the committee can understand what improved and what new obligations accompany the revision.

An efficient gaming review compares the pattern of revisions across candidates. A new readiness definition applied only to the sponsor's favoured proposal suggests asymmetric treatment; the same revised definition applied to every comparable proposal may be a legitimate correction. A cost exclusion that appears in every scope is different from an exclusion that appears only after one candidate loses. Inspect the timing and distribution of the changes before inferring intent. The decisive evidence is the comparison rule and its application, rather than the length of the edit log. A suspected bias can therefore be challenged precisely without treating every late improvement as misconduct.

The committee should be willing to reject the model use when its incentives overwhelm its evidential discipline. If sponsors cannot supply comparable scope, if ratings are negotiated to achieve an outcome or if the universe is repeatedly tuned without reason, a numerical ranking may conceal more than it reveals. The remedy can be a qualitative options review, a staged evidence process or a different appraisal method. Retaining a convenient score is not the overriding objective. The process should preserve a decision that can be challenged on its actual basis.

Control consequential changes to the decision architecture

Predeclare the candidate inclusion rule, criterion rubrics and weight-use statement. Review scope and evidence before accepting favourable ratings. Preserve the original scenario when a change is made, identify the proposition or mandate that changed and recompute a coherent comparison. Challenge omissions, overlap and range movement as well as individual numbers. Permit legitimate redesign with a clear version record. If the process cannot maintain comparable scope and accountable revisions, stop using the ranking as the basis for commitment and choose a more credible decision workflow.

The case for trusting sponsor expertise

Sponsors often know their proposals better than a central review team, and excessive challenge can slow useful improvements. Trust is appropriate when evidence, scope and accountability support it. The weakness is assuming that expertise also removes incentives or interpretation differences. The repair is a proportionate common review: verify consequential claims and changes, preserve the source perimeter and allow the sponsor to explain legitimate revisions. A transparent process should use expertise while retaining the ability to question the supplied model inputs.

The experiment establishes a mechanism, not prevalence

The range fixture demonstrates that adding an alternative can alter scores through normalisation. It does not establish how often candidate-set manipulation occurs, whether any real sponsor intends it or whether a particular organisation has weak controls. The app cannot infer intent, compare evidence quality or certify the governance process. The challenge register is an original process design for anticipated failure modes. Actual allegations or findings require evidence outside this methodological study.

Decision implications

- A rating improvement should have a changed proposition, evidence or documented interpretation.
- Candidate admission can alter the scale even when the leader stays unchanged.
- Permit legitimate redesign while preserving consequential versions and reasons.

GOVERNANCE AND APPLICATION

16

Committees need a challengeable recommendation

A capital committee recommendation template

Turn a preference into a controlled request for action

A recommendation should name the action being requested. Prefer this proposal is incomplete when the committee could authorise a study, reserve resources, approve procurement preparation or commit to implementation. Each action has different consequences and evidence requirements. The analyst should state the commitment stage before explaining the ranking. That gives the reader a practical interpretation of the preference result and prevents a preliminary comparison from being treated as permission to spend or implement beyond its scope.

The recommendation should then state its decision basis in terms a reviewer can challenge. It identifies the charter, admissible universe, evidence perimeter, governing mandate and method. These elements need not occupy the opening sentence, but they should be available immediately after it. A committee can question whether the envelope is right, whether an option was omitted or whether the weight vector has authority. An unexplained appeal to the model makes those questions harder to ask even when the underlying analysis contains the information.

A conditional recommendation is stronger when its condition changes the action. Subject to satisfactory diligence is too broad to guide the next team. Subject to confirmation that the operating arrangement supports the scored readiness interpretation is more useful because it identifies a proposition. The record should also state what happens if the proposition fails: revise the scope, reconsider a reserve option, authorise further investigation or stop. A condition without a consequence can become ceremonial language that does not constrain commitment.

The analyst should separate evidential conditions from governance conditions. Verification of a cost perimeter is an evidence task. Adoption of a strategic mandate is a decision task. A permit or approval may be an eligibility task under its applicable process. Combining them under a single diligence label obscures who is responsible and how closure is

established. The brief should assign each condition an owner, the evidence or authority needed and the disposition of the present action while it remains unresolved.

The recommendation should also say what the model did not decide. A standalone preference can support a next-stage investigation without selecting a funded combination. A Pareto result can identify represented tradeoffs without certifying feasibility. Selected weight challenges can expose conditionality without estimating probability. These boundaries belong near the action because they determine how the recommendation may be used. They should not be left solely in a technical appendix that readers may never consult.

A completed fictional committee recommendation

The completed record below uses the paper balanced decision envelope qualitatively. It is a design illustration of a reviewable next-stage recommendation, not a transaction instruction or report of an actual committee.

A completed fictional committee recommendation		
Briefing element	Completed statement	How it can be challenged
Requested action	Authorise a bounded operating-design investigation of the reliability intervention	Is investigation the commitment currently within the committee authority?
Preference basis	Use the declared balanced mandate and common standalone performance perimeter	Does this mandate govern the present objective and stage?
Evidence condition	Confirm the operating responsibility and integration assumptions supporting readiness	What finding would require a revised rating or scope?
Eligibility condition	Resolve applicable approvals through their required processes before implementation	Does the preference brief improperly imply approval has already been obtained?
Reserve treatment	Retain other admissible proposals for mandate and compatibility analysis	Would a changed charter or failed condition make a different action preferable?
Closure rule	Return with verified findings, revised inputs where needed and a proposed disposition	Is the next meeting asked to decide something that remains genuinely open?

Original completed briefing record. Numerical support is in the frozen baseline and sensitivity exhibits; this record states the action and its governance conditions.

Interpret the amendment before changing the analysis

Dissent should be specific enough to improve future review. A participant may accept the calculations while preferring a different mandate or questioning a rating. The brief should record the alternative interpretation, the action it supports and the condition under which it becomes decisive. This gives the committee a usable minority record. A generic statement that the proposal received mixed views loses the content of the disagreement. It cannot help the next reviewer distinguish a resolved issue from a concern that remains material.

The analyst should not present dissent as a defect to be hidden. A well-defined alternative scenario can expose the limits of the adopted mandate and make the recommendation more credible. The committee may still decide to proceed. The record then explains that it accepts the tradeoff while retaining the dissenting condition for monitoring. That is different from claiming consensus after a vote or after averaging weights. Agreement about the action can coexist with disagreement about the rationale, provided the distinction is visible.

Consider an amendment to the fictional recommendation: the committee permits an operating-responsibility investigation but excludes procurement preparation. The preferred intervention and the balanced mandate may remain unchanged. What changes is the authorised work. The analyst should retain the original request and record the narrower investigation separately, with its permitted evidence work and return condition. Describing this as approval of the preferred proposal would erase the exclusion. The operating team could then read a preference result as permission to prepare a purchase that the committee never authorised. This is a proposed illustration of a record, not an account of a committee that has actually met.

A different amendment changes the governing mandate rather than the permitted stage. Suppose the committee asks for a resilience-led comparison while keeping the same raw observations. The balanced result remains a reproducible description of its original question; it is no longer sufficient evidence for the amended question. Preserve the original input and output, create the expressly changed scenario, and explain the contribution changes before writing the revised recommendation. If a committee instead requests a funded combination, changing weights is insufficient: the admissible packages, resource constraint and interactions

need a representation. The minutes should identify which question was amended so that later readers do not attribute a changed recommendation to new operating evidence that was never supplied.

The minority record becomes useful when it specifies its relationship to such an amendment. A participant can accept the narrower investigation while opposing implementation until the integration dependency is evidenced. The adopted action and that dissent are compatible. Record the unresolved proposition and the later stage it constrains, rather than counting the participant as either an unconditional supporter or an opponent of all work. If the dependency is subsequently supported, the record supplies a reason to revisit the implementation objection; it does not turn the earlier investigation into retroactive implementation approval. This distinction preserves the minority's actual reasoning while allowing the present bounded action to proceed under the committee's real authority.

Distinguish deferral, rejection and the adopted commitment

A deferral and a rejection also carry different propositions. Deferring an implementation request because operating responsibility is unresolved leaves an evidence task: determine who can perform the required integration work and under which arrangement. Rejecting the request because the admissible scope excludes that intervention closes the present option under the charter. Collecting the same evidence cannot reverse the latter disposition without a changed charter or proposal. A low preference rank settles neither question. The briefing should identify the cause of the negative action and the event capable of reopening it. This gives a sponsor a specific response instead of inviting an expanding dossier around a closed question.

The final record should let a later reviewer compare the requested and adopted commitments without reconstructing the meeting from model files. In the narrower-action illustration, retain the requested operating-design investigation, the explicit exclusion of procurement preparation, the remaining integration question and the proposed return with verified responsibility evidence. Link the retained calculation to the edition used at the meeting; link any amended mandate to its separate scenario. A subsequent implementation recommendation must then state its own authority and evidence. The archive can distinguish a faithful narrower decision, a changed analytical question and unauthorised scope

expansion. Numerical replication alone cannot establish which of these occurred, because the model records a comparison rather than the institution's adopted commitment.

How a recommendation can fail and what to repair

This original exhibit gives a reviewer a direct route from a weak briefing element to a repair. It prevents a single generic demand for more detail from obscuring different problems.

How a recommendation can fail and what to repair		
Weak element	Why it fails	Concrete repair
No commitment stage	A preliminary preference can be mistaken for implementation approval	Name the action, scope and authority requested now
Generic diligence condition	No proposition or disposition governs closure	Specify the evidence owner, finding and action under failure
Unexplained governing vector	A calculation can conceal an unauthorised tradeoff	Link the mandate to its decision rights and rationale
Leader presented as optimal portfolio	Funding and interactions were never represented	State the standalone scope and commission package selection if required
Dissent summarised as opposition	The substantive alternative cannot inform later review	Record the alternative scenario and triggering condition
Output attached as minutes	The action actually adopted remains ambiguous	Write the adopted scope, conditions and authority separately

Original recommendation challenge map. Each repair changes the reviewability of the action rather than merely increasing the report length.

Write a recommendation that can be accepted, rejected or revised

Name the present action and decision authority. State the governing charter, evidence perimeter and mandate. Explain the preference with raw and contribution evidence, then show the relevant challenges. Assign each unresolved condition an owner, closure criterion and consequence. Preserve substantive dissent and method limits. Finish with the next reviewable action. After the committee decides, record what it actually adopted separately from the model output so the archive preserves both the analysis and the commitment.

The case for a decisive unqualified brief

A confident brief can help a committee act when analysis has become diffuse. Decisiveness is useful when it names a clear authorised action. It becomes misleading when it erases uncertainty or treats a preference calculation as a complete commitment case. The repair is a direct conditional request: state what should happen now and which material conditions govern it. This can be concise and decisive while retaining a falsifiable basis. A qualification that changes the action strengthens the brief; an indiscriminate list of remote caveats does not.

The completed record is a process illustration

The fictional briefing is not evidence of an actual committee, approval or sponsor agreement. The app does not draft authoritative minutes, verify conditions or enforce decision rights. Its exports support inspection and replication of numerical scenarios. A real recommendation requires an accountable author and institution-specific authority. The paper demonstrates how to connect a conditional preference to a reviewable action without claiming that the model itself authorises commitment.

Decision implications

- Name the action and stage before presenting the preference.
- A condition should have an owner, closure criterion and consequence.
- Record the committee action separately from the computed output.

GOVERNANCE AND APPLICATION

17 A pilot can buy decision flexibility

A staged-investment decision protocol

Make learning milestones capable of changing a later action

A pilot is useful when a bounded commitment can resolve a proposition that matters to a later action. It should have a question, a credible observation method and a decision that remains open after the result. A small implementation labelled pilot does not automatically create learning. If the organisation has already committed to scale regardless of the findings, the activity may be a rollout stage rather than a choice experiment. That purpose can still be legitimate, but the recommendation should describe it accurately.

The learning question should be narrower than the ambition of the whole programme. A pilot may test whether the operating team can respond to an alert, whether an interface supports a specified workflow or whether users adopt a process under stated conditions. It cannot necessarily

establish the full long-term impact of a scaled intervention. The team should identify which part of the outcome mechanism the bounded activity examines and which parts remain outside it. This makes a favourable result informative without exaggerating what it proves.

The committee should connect the question to the disputed rating or condition. If readiness is uncertain because no response workflow has been demonstrated, the pilot should examine that capability. If impact is uncertain because the causal mechanism is unclear, a process demonstration may not resolve it. A pilot designed around easy-to-measure activity can leave the consequential uncertainty untouched. The evidence record should state which input interpretation or commitment condition would change under different findings.

A useful pilot design includes an alternative explanation challenge. Improvement during the activity might reflect seasonal conditions, additional temporary staff or changes occurring elsewhere rather than the proposed intervention. The appropriate evaluation design depends on the question and context. The team should establish what comparison or process evidence is feasible and what causal conclusion it can support. The paired preference app does not perform evaluation analysis; it receives the resulting performance interpretation only after the relevant evidence process has been reviewed.

Evidence and scope

Evaluation planning should inform intervention design and subsequent learning.

Evidence: [Magenta Book: Central Government guidance on evaluation](#)

A completed fictional learning protocol

The original protocol examines whether an operating response capability supports a later readiness interpretation. It contains no claimed measured effect or predetermined scale decision.

A completed fictional learning protocol		
Protocol element	Completed design statement	Decision consequence
Question	Can the accountable team receive, interpret and act on a defined operational signal within the intended workflow?	A favourable result can support that capability component of readiness; it does not establish programme impact
Scope	A bounded workflow demonstration using the proposed roles and interfaces	The finding applies to the demonstrated configuration and conditions
Alternative explanations	Record temporary support, concurrent changes and departures from normal operation	Do not transfer a favourable finding to unsupported routine conditions
Adverse finding	Identify whether the failure concerns data, interface, training or operating responsibility	Redesign the failed component and reconsider the readiness interpretation
Scale decision	Return to the committee with the capability evidence and remaining programme requirements	Scaling remains a separate authorised commitment
Exit	Stop the bounded activity when its question is answered or the protocol cannot produce useful evidence	Avoid indefinite operation under the label of learning

Original completed learning design. The committee must define an appropriate evaluation method for its actual operating context.

A milestone should describe a finding or verified capability, rather than merely an activity completed. A report delivered, a workshop held or a system installed can be necessary work without resolving the learning question. The committee should ask what conclusion the milestone permits and what remains uncertain. This prevents a staged plan from progressing automatically through administrative completion while the relevant decision proposition remains unsupported. Progress should be linked to evidence that changes the readiness of the next commitment.

The protocol should specify favourable, adverse and inconclusive findings. An inconclusive result is not a disguised success. It may mean the activity lacked sufficient variation, the measurement failed or the context did not expose the condition of interest. The team should decide whether redesign, a different method or stopping is appropriate. Predeclaring this

disposition reduces the temptation to interpret any activity as evidence for continuation. Learning includes discovering that the proposed experiment cannot answer the question.

A credible stopping rule protects the option to learn. It identifies conditions under which the organisation will stop, redesign or decline scaling and names who can make that decision. The sponsor should understand the rule before the pilot begins. If cancellation is institutionally impossible or the initial commitment creates unavoidable continuation obligations, the staging claim needs qualification. The organisation may still proceed, but should recognise the exposure it has accepted rather than treating the next stage as freely optional.

The pilot should use a scope representative enough for the proposition being tested. A demonstration with exceptional staff, simplified interfaces or unusually supportive conditions may reveal technical possibility while saying little about routine capability. That can be useful if the question is technical feasibility. It cannot automatically support readiness for normal operations. The record should distinguish the environment tested from the environment of the proposed implementation. A transfer plan needs evidence about the differences that matter.

The cost of learning belongs in the decision. A study or pilot consumes resources, delays some commitments and may create temporary operating burden. Its value depends on whether the information can improve a later choice enough to justify those costs under the relevant decision framework. This paper does not calculate expected information value for the portfolio fixtures. It asks the team to identify the decision that learning changes, the plausible findings and the burden of obtaining them. A vague promise to learn is insufficient to justify an indefinite programme.

Information can also arrive too late to matter. If procurement or implementation becomes irreversible before the pilot result is available, the activity may not inform the claimed decision. The team should align the learning milestone with the commitment sequence. A useful schedule shows when the choice remains open and what resources must be reserved to preserve it. The pilot recommendation should avoid claiming a staged option when the organisation has already crossed the consequential commitment boundary.

The ranking can help choose what to investigate when its decision envelope is designed for that purpose. A development allocation may emphasise the importance of resolving an uncertainty and the feasibility of learning, rather than the same readiness rubric used for

implementation. The canonical app has fixed criteria and does not calculate an optimal learning programme. The team should therefore avoid pretending that an implementation ranking automatically selects pilots. It can author a clearly bounded preference scenario if the existing attributes are meaningful, or use a more suitable learning method.

A favourable pilot result should update a proposition, not simply boost every rating. Demonstrating a response workflow may support readiness while leaving strategic impact and execution risk partly unresolved. The analyst should identify the specific rating interpretation changed and retain the other assumptions unless new evidence justifies revision. This discipline makes the recomputed preference informative. A general uplift after a successful activity can double-count optimism and make it impossible to understand which evidence drove the new recommendation.

An adverse result can improve the decision even if it lowers the proposal's score. It may prevent a larger unsuitable commitment, identify a necessary redesign or clarify that an alternative is preferable. The sponsor should not be rewarded only for positive findings. A credible process values information that changes the action appropriately. The committee can preserve the pilot record as evidence of disciplined learning rather than treating a stopped proposal as a failure of analysis or effort.

At pilot close, compare the finding with the adverse explanation specified before the activity. A smooth demonstration may show that the process works with intensive specialist assistance while leaving ordinary support unresolved. Record success for the assisted condition and keep the ordinary-support branch open. If the protocol expected unsupported operation, that result fails the intended test even though the visible output is excellent. If assistance was part of the intended operating design, its continuing availability and commitment belong in the next scope. This comparison makes the pilot informative without rewriting its success criterion to match the outcome.

Transfer from a pilot by inspecting the conditions that produced its result. A small trial may exclude peak duty, difficult interfaces or the proposed operator population. The next commitment should target whichever omitted condition could defeat the mechanism, rather than repeat the same demonstration at greater expense. If the omission is an integration interface, test that interface; if it is sustained duty, investigate performance over the relevant duty; if it is operator capability, examine ordinary staffing. A successful trial can justify these next tests while remaining insufficient for full implementation. The staging sequence should progressively remove distinct failure explanations.

Approve a learning commitment with an open later choice

Name the proposition and later action that the pilot can change. Define a bounded scope and an evaluation approach capable of examining the proposition. Predeclare favourable, adverse and inconclusive dispositions, along with a credible exit and responsible decision maker. Align the finding with the commitment sequence so the later choice remains open. Update only the input interpretations supported by the findings, recompute the relevant scenario and retain the transfer limits. Authorise the next stage on the evidence obtained rather than on activity completion alone.

The case for implementing and learning simultaneously

Some interventions can be improved during implementation, and a separate pilot may delay useful service. That approach can be reasonable when the commitment is sufficiently reversible, the learning question remains observable and the organisation can change course. Its weakness is calling an effectively irreversible rollout a learning option. The repair is to identify the commitment boundary, residual exposure and decisions that can still change. Where a bounded experiment can resolve a consequential uncertainty before that boundary, a separate stage may be justified.

No empirical effect or information value is estimated

The completed protocol is fictional and contains no measured pilot result. Portfolio Compass does not design experiments, estimate causal effects, calculate sample sizes or value information. A rating update requires an appropriate evidence process outside the app. The chapter supplies a decision architecture for learning, with scope and stopping conditions. It cannot certify that a proposed pilot will resolve the uncertainty or that its information benefits exceed its costs.

Decision implications

- A pilot needs a proposition and a decision that remains open.
- An inconclusive or adverse finding requires a predeclared disposition.
- Update the specific interpretation supported by evidence, rather than every rating.

GOVERNANCE AND APPLICATION

18 Performance review must preserve evaluation logic

A portfolio learning and change-control record

Keep the original logic while responding to material change

A decision model has a controlled life after the meeting. New evidence, changed objectives or altered constraints can make its recommendation less applicable. The organisation should decide in advance which changes require review and who owns that review. Without triggers, the model may be reused long after its scope has changed. With indiscriminate review, every minor update can reopen the whole debate. A useful process focuses on changes that can alter eligibility, preference, feasible action or the conditions of the adopted commitment.

The original logic should remain available. Archive the charter, inputs, governing vector, model version, conditional recommendation and actual action adopted. The archive is not merely a collection of output screenshots. It needs the assumptions and reasons that explain the action. A later reviewer should be able to distinguish a calculation error from an unsupported estimate, a changed mandate or a failure to satisfy a condition. Those distinctions guide repair. Recalculating the current numbers alone cannot reconstruct what the earlier committee knew and authorised.

The review should first ask whether the decision is still the same. A change from investigating an option to implementing a programme can require a different evidence standard and method. A new funding envelope can make package selection relevant. A changed service objective can alter the candidate universe. If the envelope has changed, the organisation should revise the charter rather than simply update values in the old model. Numerical continuity does not guarantee continuity of the decision being made.

Performance evidence should be updated proposition by proposition. If a verified operating plan resolves part of readiness, record the corresponding interpretation and remaining uncertainty. If a cost estimate includes a newly recognised enabler, identify the scope change before comparing it with other candidates. The team should avoid treating every revised value

as comparable with the old one. A changed perimeter may require reassessment of the whole set. The review record should say whether the change concerns evidence within a fixed definition or a revision to the definition itself.

Evidence and scope

Assumptions, alternative evaluation and later verification should be explicit.

Evidence: Circular A-94: Guidelines and Discount Rates for Benefit-Cost Analysis of Federal Programs

A completed review-trigger register

The original fictional register ties each trigger to a specific review rather than reopening the model for every event. No event is claimed to have occurred.

A completed review-trigger register			
Trigger	What may change	Review owner	Required response
Operating design verified or fails	Readiness interpretation and an implementation condition	Operating owner and committee analyst	Update the supported proposition, recompute the scenario and apply the closure disposition
Cost perimeter expands	Raw cost comparability and funding feasibility	Cost-estimate owner	Align the perimeter across candidates before treating score changes as performance changes
Strategic envelope revised	Objective, weights and candidate universe	Authorised committee	Reissue the charter and adopt a mandate suitable for the new decision
Enabler becomes unavailable	Eligibility, scope or dependency feasibility	Programme owner	Reconsider the affected action and reserve options outside standalone ranking
New credible candidate admitted	Observed ranges and comparison set	Option-register owner	Apply the inclusion rule and recompute the entire coherent set
Model version changes	Computational semantics or validation	Model maintainer and analyst	Reproduce archived fixtures and disclose any changed result before using the new version

Original change-control exhibit. The response concerns the decision architecture and evidence record, not an automated action by the app.

A retrospective review should compare propositions with the evidence later obtained, without converting the old score into a forecast it never was. The model did not predict a financial return or a probability of success. It expressed preference under supplied ratings. A useful review asks whether the readiness interpretation was supportable, whether the claimed impact mechanism operated and whether the scope omitted material commitments. Those findings can improve rubrics and evidence standards. They should not be described as calibration of a probability that the model did not estimate.

Outcome review needs attention to counterfactual limits. An implemented proposal's observed outcome may depend on context, execution choices and other interventions. A favourable outcome does not prove that the original selection was uniquely optimal, and an adverse outcome does not automatically prove that the preference model was wrong. The organisation should evaluate the relevant mechanism with an appropriate method and distinguish selection quality from delivery performance. The archive helps identify what was assumed and what actually changed after approval.

Process review asks whether the decision was made as intended. Were eligibility conditions resolved? Was the governing mandate authorised? Did the sponsor implement the scope that was scored? Were review triggers acted on? These questions can be answered without claiming a causal estimate of programme impact. A process failure may explain why a plausible recommendation did not translate into the intended action. Repair then concerns governance or execution rather than recalibrating the weights to fit the later outcome.

The team should avoid hindsight rewriting. Once a result is known, it is easy to describe the winning or failed proposal as obviously superior or inferior. Preserve the evidence available at the decision date and the uncertainty that remained. A fair review asks whether the action was defensible under that information and authority, while still learning from later findings. This does not excuse avoidable mistakes. It prevents the organisation from confusing bad luck, new context and weak original reasoning.

A changed weight vector requires a changed rationale. The committee may legitimately value readiness more after a service obligation becomes urgent, or impact more after adopting a strategic expansion envelope. The record should identify the event and authority behind that revision. It should not describe a vector adjusted to match a realised outcome as

evidence that the new weights were always correct. Preferences are governed judgements, not parameters to be retrospectively fitted without a decision mandate.

Model maintenance also needs a distinction between corrections and enhancements. A correction may repair a validation error or a numerical defect while preserving intended semantics. An enhancement may change normalisation, sensitivity or the criteria represented. Archived scenarios should be recomputed and compared under the new version, with any changed interpretation disclosed. The platform's version validation prevents importing an unsupported version silently. The organisation still needs a release record explaining why the change was made and which earlier decisions require reconsideration.

A new candidate is not merely another row. Under current-set normalisation it can change every transformed range. The review should therefore generate a new scenario rather than append the candidate to a briefing containing old scores. The original set remains a valid record of the earlier comparison. The new set answers an updated question. Keeping both with their dates, reasons and scope prevents a reviewer from mistaking scale movement for a change in individual performance.

Monitoring can reveal that the score selected a proxy rather than the intended outcome. A programme may increase the number of functioning sensors while leaving operational decisions unchanged because no team uses the readings. The installation count then supports delivery, while the benefit mechanism remains unestablished. Trace the chain from installed capability to use and from use to the valued service before revising impact. This diagnostic distinguishes a failed intervention, a failed adoption route and a failed indicator. Assigning a benefits owner is useful only when that owner can investigate those links rather than report the convenient quantity repeatedly.

A changed ranking after commitment creates a switching question with a different reference action. Continuing the installed intervention may avoid transition disruption, while switching may require new integration, stranded work or retraining. The original acquisition comparison does not contain those consequences. First identify what can still be changed and which costs or obligations belong to each current route; then select a method capable of comparing continuation, adjustment and replacement. A reserve option's earlier preference score is relevant history, not a sufficient switching justification. Continuing despite a lower present score can be rational when the new alternatives face consequences absent from the old choice.

Select review events by the causal position of the changed information. A verified cost exclusion can change the commitment perimeter immediately; a new candidate can change range endpoints even without altering existing raw values; a minor operating observation may require a longer comparable record before changing an impact judgement. These events should not all trigger the same mechanical refresh. Identify the affected proposition and dependent comparison first. An unchanged mandate with stable evidence needs a status check, while a changed universe needs scale interpretation and a changed objective needs a new mandate. This triage reduces repeated calculation that adds no decision information.

Historical reconstruction should preserve two views of a changed proposal. The contemporaneous view asks whether the original action followed reasonably from the evidence and preferences available then. The present view asks what action is justified now. Using today's evidence to overwrite the earlier fixture destroys the first view; continuing to apply the earlier fixture as current advice destroys the second. Keep the editions and connect them with a proposition-level difference explanation. Exact input and output records support reconstruction, while the explanation identifies whether the change concerns performance, preference, eligibility or the engine. Neither view alone supplies the whole learning account.

Review the proposition and the action, not only the score

Preserve the original charter, evidence, vector, engine version and adopted action. Apply predeclared triggers to determine whether the present decision remains within that envelope. Classify the change as evidence, definition, mandate, feasibility or computation. Update a coherent scenario and explain its consequence for the action. Review outcomes with an appropriate evaluation method and processes against their conditions. Retain the earlier edition so hindsight cannot rewrite the original basis. Close the review with a disposition, owner and next material trigger.

The case for a continuously refreshed ranking

A frequently updated ranking can make new evidence visible and support an active development pipeline. It is useful when changes have defined owners and consistent scope. Its weakness is that automatic refresh can obscure changed ranges, mandates or commitment stages while making the result look like a continuous measure. The repair is to version consequential changes and show the reason for each new scenario. Refresh should improve applicability, rather than create the appearance of new analysis without a changed proposition.

The model does not monitor outcomes or enforce reviews

Use accountable institutional roles and appropriate evaluation methods to gather operating evidence, assess causal outcomes, enforce review triggers and verify conditions. Portfolio Compass only recomputes the supplied scenarios; the fictional review register assigns those duties without performing them. Keep the retrospective score within its original meaning as conditional preference. It was never a probability or monetary prediction, so later review cannot treat it as a calibrated forecast. Reconsider the action through the relevant evidence and mandate when either has changed.

Decision implications

- Review triggers should connect material change to an accountable response.
- Later outcomes do not turn preference points into a forecast retroactively.
- Keep the original edition and distinguish evidence updates from changed definitions or mandates.

GOVERNANCE AND APPLICATION

19 When transparent scoring is the wrong instrument

A method-selection boundary

Use aggregation only when its operations fit the action

A transparent weighted model is useful when the decision involves several meaningful criteria, admissible alternatives and compensable performance tradeoffs. It is not the default answer to every capital question. The team should select the method after defining the action, evidence and constraints. An organisation that begins with a scoring template can force unlike problems into the same form. Method choice should explain what the analysis needs to represent and which operations are justified by the available information.

Some decisions do not need aggregation. If a candidate is no worse on every relevant represented attribute and better on at least one, a dominance argument may resolve the comparison subject to omitted considerations and evidence quality. If only one admissible option can meet a binding service requirement, feasibility and affordability may be the central questions. A weighted score can add presentation complexity without changing the action. The analyst should be willing to use a simpler argument when it is more direct and adequately supported.

Other decisions need more structure than the additive app offers. A funded programme with dependencies and shared benefits requires package or constrained selection analysis. A financial investment case requires cashflows and an appropriate economic framework. A causal impact question requires evaluation. A dispute about non-compensable rights or authority requires the relevant governance process. The preference model can remain a supporting exhibit, but it should not be asked to perform operations outside its design. The method boundary should be stated before commitment, not discovered after the score is challenged.

A qualitative options review can be sufficient at an early stage. The team can compare outcome mechanisms, evidence gaps, scope and eligibility without inventing precise ratings. This is particularly useful when the main work is clarifying alternatives or establishing a common rubric. A premature numerical model can create apparent comparability before the concepts are ready. The analyst should consider whether the present action requires an ordering or simply a set of development dispositions. A structured qualitative register can be deep and reviewable without a total score.

Evidence and scope

Project preparation distinguishes value-for-money assessment from affordability.

Evidence: EPEC Guide to Public-Private Partnerships

Match the decision problem to an analytical form

This original method map names the representation needed for the decision. It does not imply that a method is valid merely because its label appears in the table.

Match the decision problem to an analytical form			
Decision problem	Representation needed	Suitable analytical direction	Role of standalone MCDA
Compare admissible independent alternatives with several compensable attributes	Common criteria, meaningful scales and authorised tradeoffs	Transparent multi-criteria preference comparison	Core preference explanation when its assumptions hold
Choose a funded programme	Budget, timing, indivisibility and interactions	Constrained package or allocation comparison	Screening input; scores are not automatic package utility
Assess financial viability	Scoped cashflows, financing and relevant risk assumptions	Financial and economic appraisal	Optional non-monetary perspective beside the appraisal
Resolve an early option definition	Mechanisms, scope, evidence gaps and conditions	Structured qualitative options review	Defer numerical scoring until the rubrics are meaningful
Estimate an intervention effect	Outcome evidence and a credible causal comparison	Evaluation design and analysis	Receives reviewed performance interpretations; does not estimate effects
Resolve a binding condition or authority dispute	Applicable requirements and decision rights	Required assessment or governance process	Cannot compensate for the unresolved condition

Original method-selection exhibit. Each method still requires credible evidence, scope and an authorised decision rule.

Urgency changes the decision process, but it does not make unsupported numbers informative. In an emergency, the team may need a rapid feasibility screen, a reversible action or a rule based on a binding requirement. The appropriate method should reflect the time available and the consequence of delay. A concise qualitative comparison can be more defensible than a rushed vector whose criteria and scales nobody understands. The organisation should record the limited evidence and plan a subsequent review when time permits.

The burden of analysis should be proportionate. A low-consequence reversible choice may justify a short transparent screen, while an irreversible strategic commitment warrants more evidence and challenge. Proportionality does not mean that a small decision should use arbitrary ratings. It means the team can focus on the conditions that matter to the

present action. An elaborate framework can be wasteful if it does not improve that action. A simple method can be inadequate if it conceals consequential interactions or uncertainty.

Cost-effectiveness analysis can be useful when alternatives pursue a common outcome that can be represented consistently. Its logic differs from mixing several transformed attributes into a preference score. A cost-effectiveness ratio should have a meaningful outcome and cost perimeter. If outcomes differ substantially or include non-compensable conditions, a single ratio may not resolve the decision. The committee should not create a pseudo-efficiency ratio by dividing the MCDA score by cost. That operation lacks the interpretation established for a genuine outcome measure.

Cost-benefit analysis requires another set of judgements, including which consequences can be valued monetarily and how timing and uncertainty are treated. Its ability to express results in monetary terms can support comparison, but it does not remove the need for evidence, scope and distributional or institutional judgement. This paper does not supply a complete cost-benefit model for the portfolio fixtures. It recognises that financial and economic questions may require one. A preference score should not be used as a shortcut around the missing appraisal.

When the concern is excessive compensation or a need for preference thresholds, investigate an outranking or non-compensatory rule that represents it. Explain that method's assumptions, interpretation and parameter challenge route before using it. Changing the rule does not by itself establish legitimate authority or adequate evidence. The paired app remains an additive comparison only, and this paper supplies no computed result from the alternative methods. The next task is to design and review the required representation, with no presumption that it selects a particular proposal.

Deliberation can be essential when values or authority remain contested. A numerical model can help participants see tradeoffs, but it cannot substitute for the process through which the institution adopts them. The analyst should consider whether the present need is to calculate a consequence or to establish the governing decision principle. A facilitated discussion with separate scenarios may be the appropriate work. Calling the resulting disagreement a model error misses the actual problem and encourages fruitless numerical adjustment.

Method choice should consider evidence maturity. Cardinal-looking numbers are weak when the source only supports broad categories. Linear transformation and weighted addition assume that the represented

intervals have a preference interpretation adequate to the decision. If the team cannot defend that interpretation, a qualitative or differently structured method may be preferable. The convenience of entering a rating is not evidence that the rating supports precise aggregation. The committee should understand the strength of the representation it is accepting.

A useful check against method shopping is to state the omitted mechanism without naming the desired winner. If the objection is that two options share an indivisible resource, the need for package analysis remains whichever proposal ranks first. If the objection is a non-compensable service condition, the gate applies equally to every option. If the objection disappears when the favoured candidate leads, it has not yet established a method-fit problem. Retain the original result and test the proposed replacement rule against the same decision requirements. This symmetry test disciplines method revision without preventing a justified correction after a limitation is discovered.

A hybrid process can be useful if its components have clear roles. Eligibility screening may precede financial appraisal, preference comparison and package selection. The team should identify how information passes between them and which conclusions each component supports. A hybrid becomes confusing when the same concern is counted repeatedly or when an output is transferred without its meaning. For example, a preference score should not become a financial benefit input simply because a later model needs a scalar. The handoffs need explicit semantic boundaries.

Method selection can end in a concrete representation brief. State the decision object, the disputed mechanism and the required output. For a standalone development comparison, retain distinct criteria and examine preference tradeoffs. For implementation under shared resources, require feasible packages and explicit resource use. For a mandatory service outcome, compare credible ways of meeting that outcome before discussing discretionary advantages. For a contested objective, resolve the value dispute through the relevant deliberation. These branches request different analytical work. Their specificity helps the recipient select a suitable instrument while avoiding an unsupported promise that a more elaborate model will automatically settle the choice.

Select the method after identifying the required representation

Define the action, admissible alternatives and binding conditions. Ask whether dominance or feasibility already resolves the choice. Identify whether the remaining question concerns compensable preference, funded combinations, financial value, causal effect or authority. Check that the evidence supports the proposed scale and operations. Assign each analytical component a clear role and handoff. Record the method-fit rationale before inspecting a preferred outcome, and revise it only for a disclosed limitation. Name the next work needed when the paired app cannot represent the requested decision.

The case for a standard scoring template

A standard template can promote consistency, lower administrative burden and help teams learn a common language. It is useful as a structured starting point. Its weakness is assuming that every problem fits the same compensatory standalone comparison. The repair is a short method-fit screen before using the template, with an explicit route to qualitative, financial, evaluation or package analysis. Standardisation should make assumptions easier to inspect, rather than eliminate the judgement required to choose an appropriate method.

Alternative methods are identified, not executed

Design and review the selected method against its own evidence and authority before applying it to these proposals. This chapter identifies fit and missing representation; it implements no cost-benefit, cost-effectiveness, outranking, constrained-optimisation or causal-evaluation study. The numerical illustrations use only the canonical additive app. A recommendation to obtain another representation leaves its result open, including which project it might select. Retain that open question while giving the next team the decision object and method-fit brief.

Decision implications

- Method choice follows the action, constraints and evidence maturity.
- Dominance or feasibility can make aggregation unnecessary.
- Hybrid analyses need semantic handoffs so an output is not used outside its meaning.

GOVERNANCE AND APPLICATION

20

The strongest case against preference models

An adversarial assessment and repair framework

Let the strongest objections define the method boundary

The strongest objection to the paper is that a transparent score can still give an unsuitable decision process the appearance of discipline. Visibility does not establish that the criteria are legitimate, that the ratings have meaningful intervals or that the people choosing weights have authority. A reproducible calculation can faithfully reproduce an unjustified architecture. The paper's thesis therefore has a condition: transparency and sensitivity improve the defensibility of a preference comparison only when its decision envelope, evidence and operations fit the action being requested.

The first challenge concerns comparability. Impact, readiness and risk ratings can look cardinal because they are numerical, while their underlying rubrics may support only broad ordinal statements. Linear normalisation and weighted addition then create precision that the evidence does not warrant. A difference between two ratings is treated as a proportional preference movement within the range. If the team cannot defend that representation, exposing the formula does not cure the problem. The appropriate repair may be a better rubric, a qualitative comparison or another method.

The second challenge concerns legitimacy. Some values should not be traded away through compensation, and some participants may have rights or decision roles that an average vector cannot represent. A clean contribution ledger can show exactly how a binding concern was outweighed without making that outcome legitimate. Eligibility and authority must therefore be established outside the sum where required. The score should operate within those conditions. It should not be used to redefine them merely because their separate treatment makes the process less convenient.

The third challenge concerns constructed universes. The model compares the candidates supplied, rather than discovering every relevant option. Sponsors and analysts can shape the universe through scope, inclusion

and timing. A high-ranked proposal may lead only because a credible alternative was never developed. The ex-ante option register reduces that discretion but does not prove completeness. The recommendation should state the search and inclusion perimeter. A ranking over a limited universe is not a global claim about what deserves capital across the organisation.

What would falsify the claimed advantage of transparent scoring

The original challenge map states conditions under which transparency is insufficient or the method should be replaced. It prevents the paper thesis from becoming a universal defence of weighted sums.

What would falsify the claimed advantage of transparent scoring			
Challenge	Failure condition	Consequence for the thesis	Required repair
Scale validity	Ratings do not support the interval interpretation used by the sum	Visible arithmetic can still create false precision	Revise the rubric or use a suitable qualitative or alternative method
Legitimacy	A binding condition or decision right is traded away	Transparency does not make compensation acceptable	Resolve eligibility and authority outside the sum
Universe	A credible relevant option is omitted without reason	The preference concerns an unjustified choice set	Reopen option development and inclusion before interpreting the ranking
Structure	Package benefits or dependencies determine the feasible action	Standalone ranking does not answer the funded decision	Use a package or constrained selection method
Evidence	Consequential ratings cannot be supported or bounded credibly	Reproducibility does not establish the decision basis	Obtain evidence, stage commitment or disclose an authorised residual uncertainty
Incentives	Architecture is repeatedly tuned to secure a preferred outcome	The audit trail records manipulation without correcting it	Restore accountable rules or stop using the score as commitment support

Original adversarial framework. A positive calculation result does not close any of these decision-fit questions.

The fourth challenge concerns meaning across scales. Observed-range normalisation can move totals when a candidate is added or a competitor's rating changes, even while an option's own raw performance stays fixed. This makes the score difficult to interpret as a stable measure of absolute value. The paper responds by treating it as a conditional preference within a frozen set and by showing the range artefact explicitly. That response

preserves a useful comparison role, but it concedes that cross-set score comparisons are generally inappropriate without a different scale architecture.

The fifth challenge concerns institutional use. A model can be technically transparent while readers see only the final ordering, a sponsor cites it as independent endorsement or a committee treats it as a substitute for judgement. Documentation cannot guarantee appropriate use. The process needs a use statement, a conditional recommendation and a record of the actual action. The paper's contribution is an architecture for that connection, not a claim that a numerical tool will automatically improve organisational behaviour.

The sixth challenge concerns the seductive simplicity of a single leader. Many strategic decisions involve incompatible objectives, different stages and feasible combinations. A single ordering can compress these differences into a format that feels decisive. The committee may then optimise the display rather than resolve the actual problem. The paper responds by distinguishing preference, eligibility, funding, learning and authority. If those distinctions make a single leader less useful, the method should yield to a different representation instead of defending the convenience of its output.

There is also a serious case for expert judgement without numerical aggregation. Experienced decision makers can recognise scope defects, operating dependencies and implausible narratives that a fixed model does not contain. For some decisions, a structured deliberation with explicit reasons may be more informative than ratings. The objection is not that judgement is error-free. It is that forcing expert interpretation into a shallow numerical architecture can discard the knowledge that matters. A transparent qualitative record can retain that knowledge while remaining challengeable.

The counterargument is that unstructured judgement also hides assumptions and can be inconsistent across proposals. A numerical model can make tradeoffs inspectable, reveal double counting and expose conditional reversals. Its value lies in those operations when their meaning is adequate, rather than in claiming objectivity. The paper does not choose arithmetic over judgement as a universal rule. It argues for an accountable representation of judgement and evidence, using a weighted model when that representation fits and recognising alternatives when it does not.

Sensitivity is vulnerable to its own performance of rigour. A report can include many scenarios while avoiding the consequential uncertainty. The selected shifts may be arbitrary, the raw ranges may be implausible or the

governing mandate may remain unresolved. The paper responds by asking each challenge to identify what changes, why the change is relevant and what action follows. A large sensitivity appendix is not itself evidence of robustness. The challenge plan needs a decision perimeter and a falsifying consequence.

The app's raw Pareto check is useful but bounded. It can flag an option that lacks a represented advantage, while omitting unmodelled features and relying on the supplied evidence. A dominated result does not prove universal inferiority. A non-dominated result does not establish affordability, permission or strategic desirability. The paper uses the check as an adversarial question rather than a certification. This avoids overstating a mathematical property of the representation as a complete conclusion about the proposal.

Use the frozen studies to inspect the engine's mandate reversals, tied scores, ignored-criterion dominance, range movement and bounded evidence changes under authored inputs. They establish mechanisms and counterexamples within that representation. To test whether the workflow improves investments, define an empirical study of actual decisions and outcomes with attention to selection and context. No comparative committee field trial is reported here. The methodological defence remains conditional until that separate evidence exists; numerical demonstrations cannot substitute for it.

Keep workflow evaluation at the level its evidence supports. Reproduction establishes a supplied preference-model behaviour; a clearer meeting record establishes a process property. Better investment outcomes require a further credible comparison with the decisions that would otherwise have been made and an account of changing conditions. One property does not establish the next. A practitioner can examine whether the raw ledger locates a dispute sooner or a recorded condition changes an action, while keeping those bounded observations separate from returns or institutional-misconduct claims. The evaluation design should identify the proposed effect before selecting its observations.

The practical test is whether the representation improves a particular decision. Can participants explain the criterion meanings, understand the tradeoffs, challenge the evidence and connect the result to an authorised action? Can a reviewer reproduce the numerical study and identify its limits? Can a changed assumption lead to a controlled revision rather than retrospective storytelling? If the answer is no, adding more decimal places or more polished charts does not fulfil the paper's thesis. The process needs repair or replacement.

Remove the weighted total as an adversarial test of the architecture. If the team can still identify the admissible alternatives, the decisive evidence gap and the next useful action, those components carry independent decision value. If the recommendation collapses because nobody can explain the leading score, the total has displaced the argument. A package problem may retain raw criteria and dominance but require another selection representation; a value dispute may retain stakeholder scenarios while withholding a collective vector. This subtraction test identifies which component is doing useful work. It supports selective adoption rather than making the complete scoring workflow an institutional prerequisite.

The strongest conclusion is therefore bounded. Transparent preference modelling and relevant sensitivity can make a compensatory standalone comparison more defensible than an opaque ranking, provided the evidence, scales, universe and mandate are adequate to the present action. When non-compensable conditions, package interactions or contested authority govern the choice, those elements require their own process. The methodology succeeds by making that boundary clear, even when the appropriate recommendation is to stop using the additive app for the requested commitment.

Apply the thesis as a conditional method test

Check whether the candidate universe is justified, the ratings support the proposed operations, compensation is legitimate and the governing mandate has authority. Confirm that standalone preference is the decision question and that consequential evidence can be supported or bounded. Use transparency and sensitivity to expose the resulting tradeoffs and reversal conditions. If a check fails, identify the missing representation and repair or replace the method. Retain useful qualitative records without defending the total merely because it has already been calculated.

The case that transparency can entrench a weak architecture

An audit trail may record every arbitrary choice while making the process look rigorous. This objection is strongest when readers equate reproducibility with legitimacy or when the model dictates a decision it cannot represent. The paper accepts that transparency is necessary for challenge, not sufficient for validity. Its response is a method-fit test with explicit exit conditions, independent evidence review and a recommendation limited to the action the model supports. Where those conditions fail, the numerical ranking should lose its commitment role.

No empirical superiority claim

Evaluate this workflow in the organisation's own context before asserting a broader performance effect. The study supplies authored illustrative fixtures and original process design, with claims about traceability and conditional numerical behaviour. It neither compares realised investment outcomes across committees nor estimates a performance advantage or empirically validates the rating rubrics. Use the stated falsifiers to test the methodological proposition and retain those unresolved empirical questions rather than treat the conclusion as universal certification of multi-criteria scoring.

Decision implications

- Transparency permits challenge but does not establish validity or legitimacy.
- The illustrative studies demonstrate mechanisms rather than empirical superiority.
- A method should lose its commitment role when its required representation fails.

PRACTICAL APPENDICES

21 Appendix: preference model specification

Reproducible canonical app methods

Publish the operations, conventions and replication boundary

The canonical computation is the exported compute function in the Portfolio Compass module. It receives a complete input object containing the relative weights and candidate alternatives, validates that object and returns a deterministic output object. The research reproducer imports this same module in the JavaScript runtime. The manuscript does not maintain a parallel finance or preference implementation in another language. Its frozen outputs are receipts of that canonical computation, while the scalar bindings identify the exact input or output field behind every printed numerical result.

The input contains the fixed criterion keys cost, impact, readiness and risk. Each alternative supplies a name and a value for every criterion. Relative weights are supplied under those same keys. The schema rejects additional fields, missing required values and malformed types. Numerical inputs must be finite and lie within the declared bounds. A small nonzero magnitude below the explicit precision floor is rejected rather than silently rounded to zero. The complete machine-readable schema is distributed with the app, so exact validation limits remain inspectable without relying on abbreviated prose.

The precision rule has a narrow purpose. It avoids accepting negligible nonzero values that can create unstable ratios or misleading representations in shared numerical routines. It is not a claim about the precision of cost estimates or rating evidence. A real estimate can be much less certain than the numerical type permits. The app's structural validation and the committee's evidence acceptance therefore remain separate. Passing the schema establishes that the computation can operate on the supplied values; it does not establish that those values are meaningful for the decision.

Candidate names must be nonempty within the allowed length and unique after the shared uniqueness check. Names organise display and export identity; they do not affect the weighted total. They can affect the order used to display equal totals because the engine uses a name comparison as its deterministic secondary sort. A reviewer should treat that secondary ordering as presentation. The substantive preference remains tied. Renaming equal-scoring candidates cannot create a justified investment priority.

The criterion transformation

For a criterion to be maximised, $v = (x \text{ minus } L) \text{ divided by } (U \text{ minus } L)$. For a criterion to be minimised, $v = (U \text{ minus } x) \text{ divided by } (U \text{ minus } L)$. Here L and U are the least and greatest supplied raw values in the current candidate set. When U equals L , every candidate receives the common midpoint value. Cost and risk use the minimising form; impact and readiness use the maximising form. These are linear current-set transformations, not externally anchored value functions.

The engine sums the supplied relative weights and rejects a vector whose total is zero. It divides each supplied weight by that total to obtain the normalised vector. Thus multiplying all raw weights by a common positive factor preserves the normalised tradeoff, provided the input

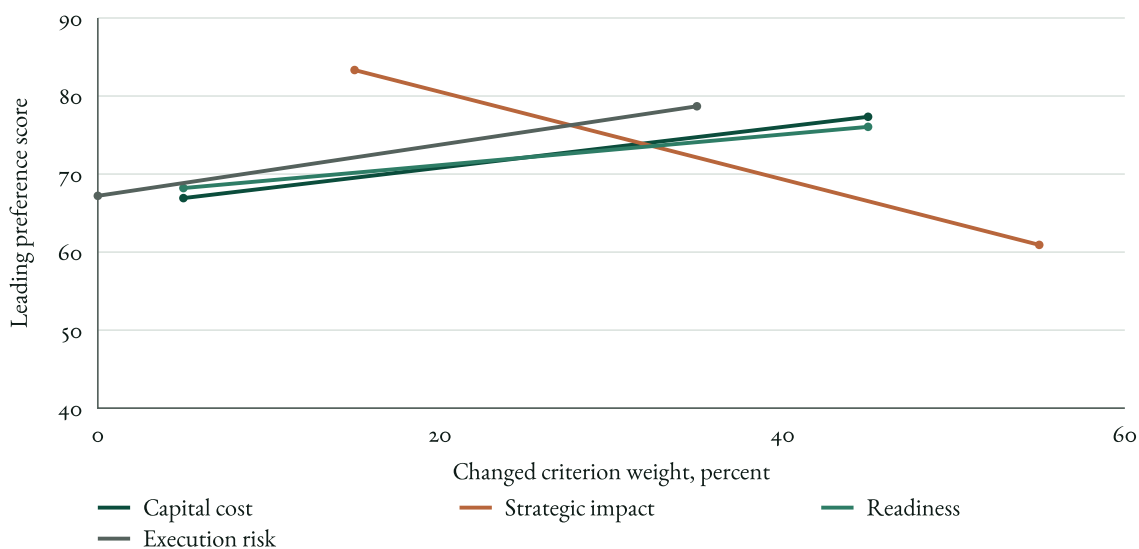
remains within the schema. The normalisation does not remove the need for an authorised elicitation rationale. It only makes the common scale of the supplied vector irrelevant to the weighted aggregation.

The transformed criterion values occupy the unit interval, with the favourable endpoint at the upper end. The score multiplies their normalised weighted sum by the app's common preference-point scale. Each criterion contribution uses the same multiplication for its own term. The output therefore contains a total and a transparent contribution ledger. The point scale is a display convention. It does not establish probability, utility in money or a proportionate measure of organisational benefit. The underlying interpretation depends on the validity of the rubrics and tradeoffs.

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

The baseline leader survives this selected challenge set

The illustrative points show the leading score under the app individual weight challenges. The leader identity remains unchanged in this fixture even though the total changes. The graphical connectors organise pairs of selected tests; they do not establish global robustness.



Original deterministic drawing from balanced-baseline canonical output pointers. Raw performances and universe remain fixed. Connectors are comparison guides between discrete tests, not an unperformed continuous sweep or a probability distribution.

Equal raw values on a criterion produce the same midpoint contribution for every candidate. That criterion does not distinguish the ordering, although it remains visible in every total. This convention avoids division by a collapsed range and makes the absence of discriminatory information explicit. It does not say that the shared performance is acceptable or

average in an external population. The common position is a numerical treatment within the present comparison. Eligibility or absolute performance requirements must be assessed separately.

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

A collapsed impact range in the illustrative dominance fixture

The impact-only example has identical impact inputs. The engine assigns equal transformed impact values and equal weighted preference totals. Raw dominance still distinguishes the candidates on the other supplied attributes.

A collapsed impact range in the illustrative dominance fixture			
Candidate	Raw impact	Transformed impact	Preference score
A costly immature system	80.00	0.50	50.00
B cheaper proven system	80.00	0.50	50.00

Canonical ignored-dominance scalar bindings. The table demonstrates the equality convention without assigning an external meaning to the common midpoint.

Raw Pareto membership is computed independently of the weights. For a candidate to be dominated, another candidate must be no worse on every raw criterion under the declared directions and strictly better on at least one. The engine checks the supplied candidates against one another and returns the names that are not dominated. It uses no preference threshold, probabilistic tolerance or inferred importance. Equality on every raw value leaves neither candidate strictly dominating the other, even if their names differ.

The distinction matters when a raw criterion has no weight. The additive score ignores that criterion by design, but the Pareto check still includes it. A candidate can therefore share a weighted total with a raw dominator. The engine retains both statements instead of inserting a hidden preference. The committee must interpret whether the zero weight expresses authorised indifference, a concern handled as a gate or a mistake in elicitation. The app’s warning and ledger make the issue reviewable without deciding it on behalf of the user.

The sensitivity routine operates on the normalised vector. For each criterion, it applies the fixed upward and downward percentage-point shift, caps the changed weight within the unit interval and rescales the others to fill the remainder in their existing proportions. When the starting vector assigns all weight to that criterion, the released remainder is distributed equally across the other criteria because no existing positive

proportion is available. That fallback is explicitly disclosed. It is a computational convention for the selected challenge, not an inferred preference.

Each sensitivity row retains the changed criterion, requested shift, resulting weight, preferred display entry, leading score, position of the original display leader and full ordering. The leader-retention summary counts how often the original display name remains first across those selected tests. It does not assess every preference vector and does not resolve tied-score substance. The close-contest appendix therefore shows the actual rows rather than interpreting the summary as confidence. The full frozen output contains the routine's details for replication.

The engine does not calculate feasibility constraints or optimise a combination. There is no budget field, package-benefit function, dependency graph, cashflow schedule or probability distribution in this input schema. Adding such concepts to a narrative around the result does not make them represented by the computation. The specification is deliberately explicit about absence because it determines what a recommendation can claim. A funding or staged programme decision needs a further method whose inputs and operations fit those relationships.

Validation and interpretation have different failure states

This original specification table identifies what the computation does when an input cannot support a current result, and what remains a human review task.

Validation and interpretation have different failure states		
Case	Computation response	Interpretation responsibility
Required numerical field blank	Reject the incomplete input and withhold current output	Resolve the evidence or author an explicit bounded assumption
All relative weights zero	Reject because no preference vector can be normalised	Elicit a meaningful authorised vector
Duplicate candidate names	Reject ambiguous identities	Assign distinct consistent candidate names
Equal raw criterion range	Use the common midpoint and issue the disclosed warning	Do not infer acceptable absolute performance
Tied preference totals	Display names in deterministic secondary order	Use an authorised rule or acknowledge indifference
Single positive criterion weight	Apply the disclosed sensitivity fallback when remainder is released	Treat the fallback as a test convention, not stakeholder judgement
Unsupported imported version	Reject rather than trusting stored output	Use a supported input scenario and recompute
Valid but unsupported evidence	Compute conditional arithmetic if supplied	Do not mistake structural validity for evidential acceptance

Implementation semantics verified against the canonical module and shared schema. These cases do not substitute for the retained automated test receipt.

Replication begins with the frozen input rather than the output summary. Import the input under the supported model version, run the canonical compute function and compare the resulting structure and scalar values with the archived output. The research audit performs this comparison and checks the binding pointers. A printed value is linked to an existing scalar, not recalculated by an author script. The author scripts only select fields, assemble explicitly written prose and organise exhibits. That separation prevents a manuscript formula from drifting away from the live app.

The claims ledger records whether a scalar comes from supplied input or computed output, its scope and the block in which it is used. Every numerical study remains explicitly illustrative. Reciprocal links let the audit check that a block’s tokens correspond to the intended claims and that claims point back to their printed uses. Unprinted output leaves

remain archived with a reason rather than being misrepresented as published findings. The ledger provides traceability, while the reader-facing exhibit explains the consequence in ordinary language.

A model-linked chart uses the same frozen scalar pointers. The deterministic plotting helper performs coordinate placement only; it does not implement preference mathematics. An adjacent plot specification records the selected outputs, drawing range and series. The audit can redraw the asset from those pointers and compare it. Lines in the selected-weight diagnostic connect discrete challenge points for visual comparison; they do not estimate the full preference boundary between them. A caption must preserve that distinction so a graphical connector is not mistaken for an unperformed sensitivity sweep.

Check the exports according to their receiving use. Reimport a scenario to verify that the supplied values and supported version survive the round trip. Inspect the CSV as a tabular record, including the neutralisation of user-controlled cells that might otherwise be interpreted as formulae. Read the printable brief to see whether a reviewer can recover the contribution explanation and its method limits. These checks ask different questions, so success in one does not substitute for another. The numerical files remain linked to the separate charter and minutes. Optional local persistence determines retention in the platform, while the institution determines retention of its authoritative decision record.

Reproducibility can be checked without claiming correctness of every supplied proposition. The automated tests cover positive, negative, boundary and invariant cases, including independent expected fixtures. Real browser journeys exercise editing, invalid input, comparison, export, import, narrow layouts and state retention. Those checks support implementation behaviour. Independent review must still challenge the manuscript's reasoning, source scope and decision limits. The edition should distinguish those evidence layers rather than using a green test receipt as a universal quality certificate.

Reproduce the canonical result before using a numerical claim

Select the named frozen input and supported model version. Run the single canonical compute source, compare the archived output and verify each printed scalar pointer. Inspect the raw set, transformation ranges, normalised vector, contributions, Pareto membership and full sensitivity rows. Check that the prose describes the same experiment and keeps its illustrative scope. Retain validation and browser receipts as implementation evidence, then apply separate source and decision review. If semantics or input scope change, create a new edition rather than silently replacing the historical basis.

The case for an independent second implementation

A second implementation can help detect a shared numerical defect when developed independently and tested against meaningful expected cases. It can also drift into a competing source of truth if manuscripts use whichever result is convenient. This project uses one canonical computation for app and paper reproduction, with independent expected fixtures and adversarial tests to challenge it. The repair for a suspected defect is to test and correct the canonical source under controlled integration, not to publish an untracked alternate calculation. Independent review remains necessary to assess the model semantics.

Exact computation is conditional on supplied meaning

The specification makes the numerical operations reproducible. It does not validate rating rubrics, estimate evidence quality, establish authority or solve portfolio allocation. Display rounding can compress close differences and does not create a substantive tie rule. Browser and automated checks have defined implementation scope. An exact scalar binding proves where a printed value came from, not that the underlying assumption is true. The recommendation must remain within the decision represented by the input.

Decision implications

- The app and research runner use the same canonical compute function.
- Raw dominance uses every fixed criterion even when a weight is zero.
- Scalar bindings and plot specifications trace results without validating supplied evidence.

PRACTICAL APPENDICES

22

Appendix: worked illustrative portfolios

Complete frozen fixtures and sensitivities

Reproduce distinct mechanisms without pooling their meanings

The worked-study appendix is a replication register, not a sample of investment outcomes. Each named study has a complete input file, canonical output file and scalar-binding file. The names identify the experiment being performed. The app schema supplies the units and structural limits, while the study metadata records the purpose, sensitivity and limitations. A reader can load a named input and recompute it using the paired current app under its supported version. The resulting calculation remains illustrative even when it reproduces exactly.

The studies deliberately vary different parts of the architecture. The balanced, impact-led and initial-capital cases hold raw performance and universe fixed while changing weights. The range-artifact case holds original performances and weights fixed while adding a dominated reference proposal. The close contest changes both the candidate profiles and the tradeoff reference to examine ties. The dominance example isolates an impact-only preference. The paired evidence cases hold the universe and vector fixed while revising the logistics readiness and risk interpretations. These are distinct experiments with different decision readings.

Read the study register as a contrast map. The mandate studies hold raw performances fixed and change preference; the evidence pair holds the mandate fixed and changes a supplied interpretation; the range experiment holds the original proposals fixed while extending the comparison set; the dominance case inspects a raw advantage that a zero-weight criterion can hide. Match each contrast to the question it can answer before comparing leaders. A preference disagreement calls for mandate work, whereas unsupported performance calls for evidence work. Counting favourable leaders across these unlike contrasts would combine different experiments into a statistic with no defined sampling or decision meaning.

Evidence and scope

Risk identification distinguishes economic, financial, legal and regulatory mechanisms.

Evidence: Identifying Risks

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

The complete frozen-study result register

This table reports the leading display entry and bounded diagnostic summaries for every currently retained study, including the supplementary range-reversal pair and positive-weight-tail diagnostic. The range-reversal mechanism is developed in the final appendix. Equal exact scores use a deterministic name ordering; the retention percentage concerns selected individual weight tests within the same fixture.

The complete frozen-study result register					
Study	Leading display entry	Score	Raw Pareto count	Retention, %	Experiment reading
balanced-baseline	Reliability architecture	72.13	4	100.00	Reference conditional preference
impact-mandate	Capacity extension	75.00	4	75.00	Different authorised tradeoff experiment
liquidity-mandate	Commercial learning	89.29	4	62.50	Cost preference without a budget ceiling
range-artifact	Reliability architecture	89.98	4	100.00	Range movement with retained original raw values
close-contest	Flexible service	50.00	3	50.00	Substantive tie and preference sensitivity
ignored-dominance	A costly immature system	50.00	1	50.00	Weighted indifference alongside raw dominance
evidence-conservative	Metering retrofit	65.83	3	75.00	Less favourable evidence interpretation

The complete frozen-study result register					
Study	Leading display entry	Score	Raw Pareto count	Retention, %	Experiment reading
evidence-optimistic	Metering retrofit	65.83	3	75.00	More favourable evidence interpretation
range-reversal-base	Lean retrofit	60.00	2	75.00	Original pair before a dominated range reference
range-reversal-extended	Deep renewal	93.33	2	100.00	Dominated addition reverses the original-pair ranking
positive-weight-tail	A	100.00	3	75.00	Positive supplied tail preserved in weight redistribution

Current canonical output scalars from the complete retained illustrative study slate. Scores are preference points; Pareto counts concern fixed raw criteria; retention is a deterministic percentage.

The balanced reference represents a fictional committee that values the cost, impact, readiness and risk swings through a declared compensatory vector. The reliability proposal leads because several represented contributions support it, rather than because it dominates every raw attribute. The capacity proposal retains the greatest represented impact while paying a preference cost for its other positions. This establishes a conditional comparison among tradeoffs. It does not establish that the reliability intervention is superior under every legitimate objective or that the larger capacity proposal should be abandoned.

The appropriate next action depends on the charter. If the present envelope concerns a bounded near-term operating investigation, the leader can support that action subject to its evidence and eligibility conditions. If the envelope concerns a funded package or strategic expansion, the standalone reference is insufficient. The study therefore provides a useful baseline and a falsifiable interpretation, not a universal project recommendation. Its contribution ledger and raw table are the first replication check before applying any sensitivity reading.

The impact-mandate study changes the preference vector while retaining every supplied raw candidate value. The capacity proposal becomes the leading entry because its represented impact receives much greater

emphasis. This is an explicit mandate reversal. It does not require the analyst to claim that capacity suddenly became cheaper, more ready or less risky. The experiment shows why a committee must authorise the tradeoff appropriate to its decision. Performance evidence alone cannot choose between these mandates.

The impact-led result remains compensatory and bounded by the same missing feasibility representation. It does not resolve permission, financing or delivery capacity. A real strategy envelope might legitimately give impact priority while still requiring gates and a credible implementation plan. The study should therefore lead to a discussion about decision scope and authority, followed by the appropriate appraisal work. It should not be used to erase the lower readiness and higher risk positions that remain in the raw record.

The initial-capital study gives cost a stronger preference while preserving the same raw candidates. The commercial-learning proposal becomes preferred. This demonstrates the consequence of a cost-oriented vector and the tradeoff it expresses. It does not impose a spending limit, model recurring obligations or compare complete cashflow profiles. A committee interested in preserving capital should identify whether its concern is an authorised preference, an affordability constraint or a timing problem. Those concerns require different representations.

The preferred low-cost proposal may also serve a different commitment stage from the capacity intervention. The charter must explain why those alternatives are legitimately compared in one envelope. If the committee is allocating learning effort rather than implementing a service, the raw impact and readiness definitions may need revision. The study is a controlled mathematical experiment; its names do not validate an actual programme design. The interpretation should remain conditional on the decision envelope that makes the vector meaningful.

The range-artifact case adds a deliberately unattractive reference proposal while retaining the original values and weights. Its raw profile expands the observed ranges. The original alternatives receive higher scores even though none of their performances improves. The baseline leader stays first. This non-reversal is part of the evidence: the fixture demonstrates scale movement, not every possible rank-reversal phenomenon. A reviewer should not infer a stronger result than the authored experiment actually produces.

The methodological consequence is that scores belong to a frozen comparison set. A higher total after candidate admission is not automatically a higher absolute value. The committee should apply its

inclusion rule and recompute a coherent set whenever the universe changes. If an externally anchored scale is appropriate, it should be designed and reviewed as a separate model rather than assumed to be present. The current app intentionally exposes current-set normalisation and its limitation.

The close contest supplies opposing profiles under equal relative weights. The service and operating-readiness proposals share the leading total, while their contributions differ. The engine displays the tied names deterministically. Selected individual shifts move the preferred entry between those profiles and can move the original display entry below the balanced candidate. The table and diagnostic identify this conditionality. They do not estimate the probability that either option is best or interpolate an analytical reversal boundary.

This study is useful for testing presentation behaviour as well as arithmetic. A briefing must not turn the literal text display into a unique substantive recommendation. The committee can acknowledge indifference, apply an authorised secondary rule or deliberate about the represented swings. The full output retains the ranking in each selected test so a reserve or staged choice can be examined. A leader-retention percentage alone would hide the tie convention and much of that decision content.

The ignored-dominance case gives preference only to impact, which is equal for both candidates. The weighted totals therefore tie. The cheaper proven system is nevertheless the sole raw Pareto entry because it is no worse on impact and better on the other represented attributes. The engine reports both results correctly under its separate rules. The apparent disagreement tests whether the committee's indifference to the ignored concerns is intended or whether those concerns should be restored to preference or handled as gates.

This is not a defect repaired by secretly adding a tie-break weight. The transparency principle requires the declared vector to govern the weighted result. A real committee might decide that cost and risk cannot be ignored, but that is a new authorised preference or eligibility decision. The original fixture should remain available as the reason for reopening it. The example also demonstrates the equal-range midpoint convention and why raw dominance can carry information that a zero-weight criterion does not contribute to the sum.

The less favourable evidence case assigns the logistics proposal low readiness and higher execution risk while retaining its supplied cost and impact. The metering proposal leads and the inspection service is the

runner-up. The labels describe the direction of authored assumptions, not their empirical probability. An actual analyst would need credible evidence supporting the bounds and a proposition that the next investigation could resolve. The fixture tests the model under those inputs without validating the logistics concept.

Its use is to identify which decision could depend on the disputed interpretation. If the present action is investigation of the leader, the logistics uncertainty does not automatically require delaying that action. If the committee needs a reserve or another compatible intervention, runner-up preference may matter. The actual action defines relevance. The fixture alone cannot infer package compatibility or the value of learning, and the report should not add those conclusions to the numerical result.

The more favourable evidence case improves only the logistics readiness and risk interpretations. The metering proposal remains first, while logistics moves ahead of the inspection service. The inspection raw values are unchanged, but its transformed positions move because the shared ranges change. The appendix preserves that distinction. A revised score is not necessarily revised performance, and a changed runner-up is not a changed leading choice. Both statements matter to honest interpretation.

The pair does not supply a confidence interval or an expected-value calculation. There are no probabilities attached to the interpretations. The correct conclusion is a bounded conditional one: the selected evidence changes affect the ordering below the leader and the normalised scores, while retaining the leading entry. Further evidence should be requested if that consequence matters to the intended action or if credible values lie outside the supplied bounds. The paper does not promise that better evidence will necessarily overturn the recommendation.

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

Complete supplied weight vectors for replication

Every retained study weight vector is shown, including the supplementary range-reversal pair and positive-weight-tail diagnostic. The current engine normalises canonical decimal values before aggregation. These vectors are deliberately authored experiments, not survey measurements or accepted institutional preferences.

Complete supplied weight vectors for replication					
Study	Cost	Impact	Readiness	Risk	
balanced-baseline	25.00	35.00	25.00	15.00	
impact-mandate	5.00	75.00	10.00	10.00	
liquidity-mandate	75.00	10.00	10.00	5.00	
range-artifact	25.00	35.00	25.00	15.00	
close-contest	25.00	25.00	25.00	25.00	
ignored-dominance	0.00	100.00	0.00	0.00	
evidence-conservative	25.00	35.00	25.00	15.00	
evidence-optimistic	25.00	35.00	25.00	15.00	
range-reversal-base	60.00	40.00	0	0	
range-reversal-extended	60	40	0	0	
positive-weight-tail	1000	1e-09	0	0	

Canonical frozen input bindings. These authored vectors are not stakeholder survey measurements or probability weights.

The complete input files are the definitive fixture records. The appendix prints the consequential parameters and result interpretations, while earlier exhibits show raw candidate tables and selected sensitivity rows. The files retain every required field and can be opened through the replication workflow. A reader should compare candidate identities rather than assuming that array position stays constant across rankings. Scalar pointers in the claims ledger identify the field selected for each printed value and prevent an author from substituting a similar-looking result from another study.

Replication should distinguish successful recomputation from acceptance of the decision. Matching an archived output verifies that the same inputs and model semantics produce the same result. It does not verify rating

evidence, legitimise a mandate or establish that the proposals are feasible. After reproducing the study, the reviewer should challenge the experiment's held-constant statement and interpretation. That sequence separates an implementation discrepancy from a methodological objection and gives each a concrete route to repair.

The studies support the conditional thesis through different challenges. They demonstrate that explicit changes to tradeoffs can reverse preference, that a zero-weight architecture can tie a raw dominator with its dominated comparator, that universe changes can move scores and that bounded evidence changes need not alter the leader. The paper's original contribution is the decision reading connecting those mechanisms to eligibility, authority, learning and funding. The frozen outputs support those particular mechanisms, while the broader process architecture remains an original proposal subject to independent review.

Read each experiment within its held-constant perimeter

Select the study that matches the methodological question. Read its purpose and complete input, identify what changes and what remains fixed, then reproduce the canonical output. Inspect the relevant raw, contribution, dominance and sensitivity records. State the result actually shown, including ties and non-reversals. Connect that result to the action only after checking charter, authority, eligibility and feasibility. Do not pool the illustrative cases as votes, probabilities or a dataset of realised investments.

The case for a single large realistic case

A detailed case can integrate scope, evidence and institutional context more naturally than several small fixtures. Its weakness in a method study is that simultaneous changes can make individual mechanisms hard to isolate, and realism can be mistaken for empirical validation. The selected fixtures separate distinct operations and retain exact inputs. A future applied case can build on them with verified project evidence and an appropriate decision process. The paper should preserve the difference between a controlled logical demonstration and a field appraisal.

A controlled fixture set is not an empirical portfolio universe

Every study is explicitly illustrative. The candidate names, raw ratings and weights are authored inputs, and the results concern the fixed additive engine only. The set does not represent actual investment outcomes or estimate institutional preferences. It contains no constrained funding solution, financial forecast or information-value calculation. Its purpose is to make distinct computational and interpretive mechanisms reproducible, with the limitations that determine how each result can be used.

A positive weight is not an absent preference

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

A supported positive preference weight should remain in the stated redistribution rule even when it is small. The positive-weight-tail diagnostic sets cost weight to 1000, impact to $1\text{e-}09$ and the other two criteria to zero. Alternatives A, B and C use capital costs 0, 0.25 and 2 USDm. Only B has impact 100; only C has readiness 100 and execution risk zero. Under the original normalised weights A leads, but applying a cost-weight shift of -20 percentage points releases weight to the sole other positive criterion, impact. The preferred alternative becomes B, at a rounded score 90.00. This illustrates the supplied preference rule, not stakeholder approval of that rule.

The revised engine normalises canonical decimal inputs, computes criterion contributions and ranks alternatives using rational arithmetic. Sensitivity distributes the remainder in proportion to the other supplied weights. It uses the documented equal-thirds fallback only when all those other supplied weights are exactly zero. Thus a positive tail is not discarded through a numeric threshold or subtraction cancellation. Exported fractions allow the case to be reproduced beyond its rounded chart values. The arithmetic does not settle whether that tiny weight is intentional, whether the criteria are measured comparably, or whether a compensatory decision is permitted. Those remain committee questions, and the diagnostic should prompt correction if the supplied preference record is wrong.

Decision implications

- The studies vary different parts of the architecture and must be read separately.
- The range and evidence cases retain their stated non-reversal results.
- Exact recomputation precedes, and does not replace, methodological acceptance.

PRACTICAL APPENDICES

23

Appendix: facilitation and committee worksheets

A usable decision session pack

Complete a session around the action that can be justified now

A substantive tie in the close-contest fixture makes the committee's next action the central question. This completed fictional facilitation session shows how to preserve disagreement and evidence limits while closing with a controlled action whose authority remains explicit. Its participants are authored process roles, not people whose preferences were measured. Follow the dispositions to see what must be resolved before selection can advance. The example supplies those completed decisions, rather than leaves an empty form for the reader to design alone.

The facilitator opens by asking the chair to state the action. The chair confirms that the present request concerns further development of an operating intervention, not approval of a funded programme. The service sponsor initially describes the goal as choosing the highest-scoring project. The facilitator reframes that statement as a method output and asks what the organisation needs to authorise now. The group agrees that the session must identify the next development action, the evidence required for implementation and the mandate governing those choices. This clarification narrows the use of the ranking without discarding it.

Present the common raw table before its weighted total. In this illustrative session, the service proposal has favourable cost and impact positions, the operating-readiness proposal has stronger readiness and risk positions, and the balanced proposal lies between them. The participants examine those swings without a preferred name determining the discussion. They retain the supplied values as illustrative and require a common rubric and perimeter for actual evidence. This sequence keeps the conversation on represented tradeoffs and the present action before the displayed ordering can become its objective.

Evidence and scope

Analytical methods should fit the evaluation question and available evidence.

Evidence: **Magenta Book Annex A: analytical methods for use within an evaluation**

The tie becomes a question for the committee

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

The illustrative model reports Flexible service and Operating readiness at the same leading score of 50.00. The facilitator explains that the first display entry follows the engine's deterministic name order. The chair records that this ordering supplies no unique substantive preference. The committee must either accept indifference under the vector, authorise another relevant rule or reconsider whether the vector fits the present development decision.

The operating owner says that the readiness swing matters because the organisation cannot support a development activity without an accountable response capability. The service sponsor says that the lower-cost proposal allows more room for learning and keeps the wider programme open. These statements are not treated as errors. The facilitator asks whether each concern belongs in eligibility, preference or a funding constraint. The owner clarifies that a minimum accountable team is a condition, while additional maturity is a preference. The sponsor clarifies that capital preservation is a preference at this stage, subject to a separately approved envelope.

The evidence reviewer assigns verification of two scope propositions raised by the group: whether a ready proposal relies on staff with other obligations, and whether a low-cost proposal omits an interface needed for learning. Keep those propositions with their verification owners and the evidence needed to support the readiness and cost interpretations. The discussion leaves the illustrative ratings unchanged rather than creates actual performance values from the exercise. A real proposal record must close those evidence questions before the committee can rely on the corresponding interpretation.

The analyst shows the selected weight challenges. The operating-oriented and cost-oriented movements produce different leaders. The facilitator asks participants to explain the tradeoff that would justify each movement in the actual envelope. The discussion reveals that they have been imagining different stages: one speaks about executing an intervention soon, while the other speaks about learning before a later implementation.

The chair decides that the charter must distinguish those stages before a governing vector is adopted. A numerical compromise is not used to hide that scope disagreement.

The committee can still make progress. It authorises preparation of comparable development scopes and an evidence acceptance record for the relevant options. It reserves implementation approval for a later meeting. This action is concrete and bounded: the team must identify operating responsibility, interface requirements and the learning question before returning. The ranked fixture remains a methodological exhibit that explains why stage and preference matter. It is not presented as an approval of the literal text display leader.

The completed session decision record

The fictional committee closes the session with a bounded preparation action and explicit conditions. No numerical preference is represented as actual stakeholder evidence.

The completed session decision record			
Session question	Recorded finding	Authorised disposition	Owner
What is the present commitment?	Develop comparable next-stage options; implementation is outside this session	Prepare scoped development records rather than approve a project	Chair and option sponsors
What does the tied score establish?	The supplied vector leaves leading profiles indifferent	Do not use alphabetical display as a substantive tie-break	Committee analyst
Which concern is a gate?	Accountable operating responsibility is required for the proposed activity	Identify and verify the responsible operating arrangement	Operating owner
Which concern is preference?	Additional maturity and capital preservation represent tradeoffs	Elicit a vector after the commitment stages are distinguished	Authorised committee with facilitator
Which evidence is unresolved?	Interface scope and staff availability may affect comparability	Prepare common-perimeter evidence records	Evidence reviewer and sponsors
What remains open?	Which development option and later implementation action should proceed	Return with the required scope, evidence and mandate for a new decision	Chair

Original completed committee worksheet. The roles and dispositions are illustrative; an actual institution must supply its authority and evidence.

The criteria worksheet is completed before the next session. Its purpose is to make each concept operational enough for consistent evidence interpretation. The group records capital cost as the burden within the stated development scope, impact as the contribution to the declared learning or service objective, readiness as the capability to perform that scope and risk as the relevant execution exposure. It also records excluded meanings. Readiness does not stand for every future programme condition, and impact does not automatically include the benefit of unfunded downstream work.

The evidence reviewer uses that worksheet to challenge overlap. A mature operating arrangement may support readiness, while the consequence of a failure belongs in the risk interpretation. The reviewer asks whether the same unresolved issue is being penalised twice without a distinct rationale. The group does not assume that any correlation proves duplication. It traces the mechanism and meaning of the criteria. Where both capture intentionally distinct aspects, it records that distinction. Where the difference is merely two labels for the same proposition, it revises the rubric before elicitation.

The option sponsors prepare scope statements using the same perimeter. Each states what the present commitment delivers, which interfaces it requires and which later decisions remain open. A sponsor cannot claim programme-scale impact while supplying only a small component cost without explaining the relationship. The review record identifies dependencies and potential exclusions. This preparation makes the next numerical comparison more meaningful because the team has addressed comparability before arguing about weights or totals.

The completed criterion interpretation record

The original record distinguishes the intended concept from a tempting substitute. It is qualitative and supplies no numerical rating rubric for actual proposals.

The completed criterion interpretation record			
Criterion	Intended interpretation	Excluded substitute	Evidence check
Capital cost	Commitments required for the stated development scope	Only the easily purchased component while enabling work is omitted	Scope, exclusions and interface obligations are consistent across candidates
Strategic impact	Contribution of the present action to the declared objective	Full future programme benefit credited to an early study	The mechanism and contingent downstream commitments are explicit
Readiness	Capability to execute the present scoped activity	Document completion or maturity of an unrelated later stage	Accountable roles and required interfaces support the scope
Execution risk	Relevant exposure during the present commitment	A generic confidence discount applied to every benefit	The risk proposition and controls are identified without duplicating readiness

Original completed rubric worksheet for a development-stage comparison. The governing committee must approve its use before actual ratings are entered.

The next facilitation session begins with the approved scopes and rubrics. Participants compare the actual performance swings those scopes produce, rather than rating criterion names in the abstract. They explain why a movement matters to the development action. If a participant cannot value readiness without assuming a particular downstream design, the facilitator records the interaction concern. The group can revise the model envelope or choose another method. It does not force a convenient weight merely to complete the form.

Record the authorised perspective, stated ranges and evidence perimeter in the vector's use statement. Its authority extends to the development comparison only; a later implementation request must reconsider applicability. The fictional record completes that procedural disposition without supplying numerical judgements from a real institution. Before elicitation, the committee therefore has a defined meaning and authority for the requested numbers, together with the conditions for reuse. If the stage or perimeter changes, reopen that use statement rather than carry the development vector forward automatically.

A dissent record preserves the alternative emphasis. The operating owner may remain concerned that the development scope creates an obligation the team cannot support. The service sponsor may remain concerned that an early implementation framing removes the value of learning. Each states the action it would prefer and the finding that would make its concern decisive. The chair can adopt the bounded preparation action while retaining both concerns for the next meeting. This is progress without a false declaration of complete consensus.

The completed dissent and reopening record

The fictional dissent statements are kept as conditional decision propositions. They are not votes or claims about a real institution.

The completed dissent and reopening record			
Perspective	Unresolved concern	Preferred conditional action	Reopening trigger
Operating accountability	The proposed activity may rely on unavailable staff	Defer execution until a credible operating arrangement is verified	The scope still lacks accountable available capability
Learning value	An implementation framing may prematurely close alternatives	Prepare a bounded learning choice before selecting implementation	The charter shifts to implementation without resolving the learning proposition
Capital preservation	Development scope may omit enabling obligations	Align the perimeter before adopting a cost-oriented preference	The revised scope adds material commitments outside the original comparison
Evidence consistency	Rubrics may treat stages differently	Reassess every candidate under the approved common definitions	A revised criterion meaning affects only a favoured proposal

Original completed minority-view worksheet. Retention supports future review without changing the authority of the adopted present action.

At closure, the facilitator reads the adopted action, conditions and owners aloud to check that participants understand the same commitment. Save the calculation separately from the decision record: the former remains the authored close-contest education fixture, and the latter the original fictional process design. In an actual session, retain the input-only scenario according to the institution's privacy and persistence choice, and keep authoritative minutes through the applicable record process. The numerical vault can preserve the calculation but cannot replace the minutes that establish the action and authority.

The fictional next team can implement the handoff through a scope comparison. Place each candidate's current activity beside the outcome credited to it and identify any downstream commitment needed to

connect them. A learning proposal credited with future programme impact needs an explicit conditional link; an implementation proposal with omitted enabling work needs a repaired commitment perimeter. Only then revisit the shared readiness and cost definitions. The preparation action remains distinct from implementation approval, but its usefulness is concrete: it can remove an inconsistent candidate comparison before elicitation. No adjustment to the weights alone can repair a row that describes a different action.

Challenge the completed fictional session by substituting the opposite procedural answer. Suppose the operating concern is treated as a low readiness score while the activity still lacks a credible accountable capability. The aggregate might then compensate for a condition the present commitment actually requires. Suppose instead that an advisory preference is labelled a gate: it could exclude a legitimate alternative before deliberation. Compare each concern's stated source and consequence with these substitutions. This identifies a classification error independently of the filled worksheet. The records support adaptation because their propositions can be tested, while the fictional roles confer no actual institutional authority.

A session should be judged by the quality of its next action, not by whether it produces a unique numerical winner. In the fictional case, the useful outcome is a bounded preparation commitment that resolves comparability and authority before selection. The tie and reversals expose why that work matters. This is a substantive decision result even though it postpones implementation preference. The committee has made progress by identifying what must be true and who must establish it before a larger commitment becomes defensible.

Close the session with shared understanding of the next action

Confirm the commitment stage, inspect raw scope and evidence, interpret ties and sensitivity without granting authority to display order, then classify concerns as gates, preference or constraints. Record the adopted bounded action, conditions and owners. Preserve substantive dissent with reopening triggers. Save the calculation separately from the authoritative decision record and hand each evidence request to an accountable role. If the session reveals incompatible stages or undefined rubrics, resolve them before eliciting the governing vector.

The case for requiring a winner at every meeting

A meeting that always ends with a winner can feel efficient and may be appropriate when the decision basis is already mature. It is less useful when the current evidence or authority cannot support that action. The repair is to require a concrete disposition rather than a unique project selection: prepare, verify, redesign, deliberate, defer or commit. The fictional session leaves with specific work and an open later decision. That is more accountable than choosing a literal text entry merely to avoid appearing indecisive.

Completed fictional records do not establish actual authority

Before adapting this tie-and-disagreement workflow, establish the institutional decision rights, evidence standards and record obligations for the actual session. The roles, dialogue and dispositions remain authored illustrations, with no real meeting, stakeholder consensus or project evidence behind them. The app neither facilitates the session nor verifies its records. Use the example to define a reviewable next action, then carry its conditions into the process that can authorise any later commitment.

Decision implications

- A tie can expose necessary scope and authority work rather than force a numerical winner.
- Completed records name propositions, dispositions and owners that a reviewer can challenge.
- The calculation and the authoritative committee action remain separate artefacts.

PRACTICAL APPENDICES

24

Appendix: source scope and a dominated-option reversal

Decide who owns eligibility and criterion endpoints before treating a preference leader as a recommendation.

The evidence boundary, stated briefly

The sources in this edition are public primary guidance and official publisher records. They serve as narrow methodological or institutional context. The manuscript's decision architecture, fictional records and fixture interpretations are original analysis. No source is treated as empirical evidence of the performance of the fictional proposals. The distinction matters because a public manual can explain a method without validating its application to a private capital decision. Each source card retains a locator and scope so the reader can identify exactly what support is being claimed.

Institutional guidance retains its original audience and authority. A public-sector appraisal manual does not automatically impose a private committee's decision rule. Corporate-governance principles do not replace the duties applicable under a particular jurisdiction. Environmental and social standards do not become a universal legal conclusion merely because the paper cites them. The manuscript uses these materials to identify questions and boundaries, while actual applicability requires the relevant institution, agreement and process. A preference score cannot perform that applicability assessment.

The methodological manuals support a similarly bounded role. They provide context for scales, tradeoffs, uncertainty and analytical selection. They do not establish that the app's fixed criteria are sufficient for every investment or that a particular vector is legitimate. The author identifies where the implemented additive comparison follows a clear operation and where it lacks representation. The conclusion is therefore conditional on decision fit. A source citation should strengthen a defined proposition, rather than create a general aura of endorsement around the whole workflow.

The publisher record for the economic-appraisal book has an especially narrow receipt. This edition retains the readable official description and uses it only for the stated distinction between economic value and investor financial profitability. It does not claim to have reviewed the full book or import unseen detailed procedures. The card's scope makes that boundary explicit. A title, summary or search result should not be expanded into a claim about contents that the author has not actually read.

The claims ledger gives numerical statements a defined epistemic status. Supplied fixture values are illustrative assumptions, while computed values are illustrative results from the canonical engine. Narrow primary guidance notes are sourced qualitative propositions. Original process reasoning is not labelled as measured evidence. These categories prevent the document from sliding between mathematical demonstration, institutional guidance and project findings. A reader should be able to identify which kind of statement is being made before deciding how it can inform a real action.

A narrow citation is not a substitute for honest intellectual attribution. If a paragraph derives substantially from a source, its use should be classified and counted accordingly rather than hidden under an original-analysis label. Conversely, a long original decision framework should not be attached wholesale to several sources as though each supports every word.

The author should separate the primary proposition from the reasoning developed around it. This makes the source audit meaningful and the reader's understanding more precise.

Source receipts, the numerical ledger and actual edition status accompany the publication. They establish different properties: access, reproducibility and recorded review. For a new capital decision, establish current institutional applicability and replace fictional assumptions with verified proposal evidence. The following controlled experiment concerns the behaviour of a preference scale; it does not claim that any capital commitment is feasible or approved.

A dominated option reverses the original leader

An additional proposal can change a recommendation even when it has no realistic claim to approval. This happens when the comparison set also defines the value scale. To isolate the mechanism, retain the original proposals, their raw criteria and the committee weights, then add a clearly dominated option. If the original ordering changes, the committee has learned something about the representation of preferences rather than the merits of a new investable proposal. The test complements the earlier range-artifact study, which moved scores without changing its leader. A non-reversal in one experiment does not establish invariance across all admissible sets.

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

The original fictional pair has Lean retrofit at USD10.00m with impact 60.00, and Deep renewal at USD20.00m with impact 90.00. Both retain readiness 50.00 and execution risk 50.00 on the app scales. Cost receives 60.00 of the declared weight total and impact 40.00; readiness and risk each receive 0. These are supplied preferences and illustrative ratings. They are not measured benefits, estimated project cash flows or a willingness-to-pay valuation.

Illustrative. On the original set, Lean retrofit leads at 60.00 preference points, while Deep renewal receives 40.00. Add Screened option at USD100.00m with impact 60.00, retaining the same readiness and risk. The original proposals and weights remain byte-for-byte unchanged. The extended set now puts Deep renewal first at 93.33, followed by Lean retrofit at 60.00. The added option receives 0.00. Its presence changes the range used to interpret the original cost difference even though it adds no non-dominated choice.

The mechanism is visible in the contribution tables. On the original pair, the cost endpoints are the costs of the eligible proposals, so the entire cost swing is their difference. The added expensive option extends only the upper cost endpoint. Lean retrofit keeps its best-cost endpoint, while the additional cost of Deep renewal now occupies a much smaller fraction of the expanded cost range. The impact range is unchanged, so the advantage of Deep renewal on impact retains its full normalised swing. Constant readiness and risk contribute nothing under the supplied zero weights. The reversal follows from this asymmetric rescaling; no improved delivery evidence or new stakeholder preference was supplied.

Raw dominance provides a useful cross-check. Both original options weakly outperform Screened option on every declared criterion and strictly outperform it on cost. The original pair remain non-dominated because their cost and impact advantages run in opposite directions. The non-dominated set is therefore unchanged, while the preference leader changes. A dominance screen and a preference score answer different questions. Keeping both visible prevents the committee from interpreting a dominated option as new strategic opportunity merely because its inclusion has changed the score of another proposal. It also exposes the consequence of including a screened-out option in endpoint construction.

The immediate response is to freeze the eligible comparison set before eliciting and approving range-dependent tradeoffs. A late proposal should trigger an explicit decision about eligibility, endpoints and preference review. Quietly adding it to the model and retaining the same verbal weight labels hides a material change in the meaning of a cost point. The committee record should distinguish a genuine change in the opportunities available from a change caused by the representation. Both may justify reconsideration, but they require different evidence and different explanations to the authority that owns the decision.

Externally anchored value functions can avoid dependence on an incidental comparison set, but choosing anchors is substantive work. They need a defended service boundary, scale shape and tradeoff interpretation. An upper cost anchor should not be selected simply to preserve a preferred leader. Anchors far outside the useful decision range can suppress a criterion just as an irrelevant alternative does. If a fixed-anchored model is appropriate, build and validate it as a separate method with its own assumptions and evidence. The current app implements endogenous min-max scaling and should continue to describe that operation accurately.

The selected weight diagnostics do not close this question. They keep the alternative set fixed while changing a weight, so they examine another uncertainty axis. A stable leader across those shifts can still be unstable to eligibility or endpoint changes. Conversely, this constructed reversal does not show how often a real committee would encounter the problem. It establishes a possibility under declared inputs. A fuller stability programme should test the plausible option universe, admissibility rules, evidence intervals and preference disputes that matter to the actual mandate, retaining coherent cases instead of inventing a probability from the number of favourable runs.

The useful committee packet contains the original pair, the expanded set, the exact criterion ranges, contributions and raw dominance results. Keep the unchanged original rows visible so a reviewer can verify the controlled comparison. State which options are eligible for commitment and which merely test the method. The recommendation should explain whether the committee accepts range-dependent tradeoffs or requires a separate anchored representation. Record the authority that can reopen the mandate and the evidence that would cause it to do so. The result is a decision about the decision method before any authority commits funds.

A dominated-option challenge that changes the leader

Completed original challenge packet for this fictional comparison. It identifies actions and observations; it is not an invented committee approval.

A dominated-option challenge that changes the leader		
Challenge	Observed in the illustration	Decision response
Original evidence changes?	Original raw values and weights remain unchanged.	Do not describe the new leader as improved project evidence.
New eligible opportunity?	The added option is dominated by both original proposals.	Apply the admissibility rule before using it to define scales.
What changed?	The upper cost endpoint extends while the impact endpoints stay fixed.	Review range-dependent tradeoffs with the decision owner.
Does raw dominance change?	The original non-dominated pair are retained.	Show dominance separately from additive preference.
What does the engine establish?	A conditional preference reversal under this constructed set.	Retain both reproducible fixtures and their contribution tables.
What remains outside scope?	Feasibility, capital-package optimisation and economic value.	Use the relevant financial, operational and programme methods.

A counterexample establishes a method boundary

This experiment diagnoses a property of the supplied comparison, not a defect in arithmetic. The model is faithfully calculating its stated min-max rule. The new option is deliberately artificial and the impact ratings have no empirical foundation. Do not interpret the leading score as a valuation or an approval. Transfer the controlled test to practice only after defining the real option universe, scale semantics and authority over eligibility.

Decision implications

- An unchanged proposal pair can change order when another option redefines its cost scale.
- Freeze eligibility and endpoints together with range-dependent preferences.
- A selected weight test does not establish stability to changes in the option universe.

PRIMARY EVIDENCE

References and source scope

wb-risk-identification

World Bank PPP Resource Center. Identifying Risks. live guidance, accessed 2026-10-10

Locator: Economic or financial; legal or regulatory

Scope: A list of possible risks supplies no empirical event frequency or project-specific probability.

<https://ppp.worldbank.org/identifying-risks>

epec-guide

European Investment Bank, European PPP Expertise Centre. EPEC Guide to Public-Private Partnerships. 2026 edition, 16 April 2026

Locator: Printed pages 9, 15, 44 and 182; physical PDF pages 17, 23, 52 and 190: lifecycle, affordability, funding and value for money.

Scope: A public-sector PPP framework is adapted here as a comparison lens; it is not a universal private investment rule.

<https://www.eib.org/files/publications/20250124-160426-epec-guide-to-ppp-en.pdf>

uk-mca

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

Locator: Manual Chapter 6, scoring, swing weighting and sensitivity; full linked PDF retained locally.

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>

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>

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>

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>

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>

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>

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

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>

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>

wb-rated-criteria

World Bank. Project Procurement: Rated Criteria. current guidance accessed 2026-10-10

Locator: Price, lifecycle cost and non-price attributes

Scope: The mandatory scope belongs to World Bank-financed procurement and is not extended to unrelated tenders.

<https://www.worldbank.org/en/about/rated-criteria>

oecd-composite

OECD, European Commission Joint Research Centre. Handbook on Constructing Composite Indicators: Methodology and User Guide. 2008

Locator: Sections 1.5 to 1.7, printed pages 27 to 35; min-max discussion, printed page 85

Scope: Country benchmarking guidance informs methodological critique; it does not validate these project scores or measure their consequences.

https://www.oecd.org/content/dam/oecd/en/publications/reports/2008/08/handbook-on-constructing-composite-indicators-methodology-and-user-guide_g1gh9301/9789264043466-en.pdf

ec-mcda-toolbox

European Commission. Better Regulation Toolbox: Methodologies for analysing impacts. 2023

Locator: Tool 62, Main features and impact matrix, printed page 552

Scope: EU impact-assessment guidance is an analytical reference, not a private investment rule. No claim of a universal superiority of MCDA is adopted.

https://commission.europa.eu/system/files/2023-09/BRT-2023-Chapter%208-Methodologies%20for%20analysing%20impacts%20in%20IAs%20evaluations%20and%20fitness%20checks_0.pdf

oecd-board-functions

OECD. G20/OECD Principles of Corporate Governance: Responsibilities of the Board. 2023

Locator: Principles V.D.1, V.D.2 and V.F; strategy, risk and access to information

Scope: The principles are policy guidance; actual duties and committee authority depend on the applicable corporate and legal framework.

https://www.oecd.org/en/publications/g20-oecd-principles-of-corporate-governance-2023_ed750b30-en/full-report/component-8.html

ifc-ps-one

International Finance Corporation. Performance Standard 1: Assessment and Management of Environmental and Social Risks and Impacts. 2012

Locator: Performance Standard 1, paragraphs 1,3,5 and 7; introduction, mitigation hierarchy, management system and risk identification

Scope: The standard retains its IFC client and contractual scope. A weighted score cannot certify compliance or authorise trading away a requirement.

<https://www.ifc.org/content/dam/ifc/doc/2023/ifc-performance-standards-2012-en.pdf>

eib-economic-appraisal

European Investment Bank. The Economic Appraisal of Investment Projects at the EIB, Second Edition. 28 March 2023

Locator: Publisher Description: economic appraisal beyond financial profitability

Scope: Only the publisher description is used here. It does not establish an economic value or appraisal result for an illustrative private proposal.

<https://www.eib.org/en/publications/20220169-the-economic-appraisal-of-investment-projects-at-the-eib>

omb-appraisal

US Office of Management and Budget. Circular A-94: Guidelines and Discount Rates for Benefit-Cost Analysis of Federal Programs. 29 October 1992, current official index link accessed 2026-10-10

Locator: Section 5.c(2) to (4): explicit assumptions, evaluation of alternatives and verification; physical PDF pages 5 and 6

Scope: Federal programme guidance retains its scope. No federal discount rate is imported into the preference model or prescribed to a private company.

<https://www.whitehouse.gov/wp-content/uploads/2025/12/a094.pdf>

hmt-business-case

HM Treasury. Guidance on Developing Business Cases for Projects and Programmes. 2026

Locator: Chapter 2, printed pages 5 to 7; Chapter 8 longlist appraisal and Chapter 10 financial case

Scope: UK and Welsh public-spending requirements retain their stated scope. The analytical distinctions are adapted without asserting a duty for every firm.

https://assets.publishing.service.gov.uk/media/6a4390675b6406df58c14006/Guidance_on_Developing_Business_Cases.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>

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: Fully authored methodological candidate with distinct canonical illustrative studies, reciprocal scalar exhibits, completed decision records and paper-local primary source receipts. No empirical project findings or acceptance claim.

2026-10-10: Added the original positive-weight-tail computational diagnostic and canonical rational preference method*i.i.o.* 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.