MathGov: A Universal Ethical Operating System for Multi-Scale Alignment

Version 5.0 (Spec-Hardened)

James McGaughran

ORCID: 0009-0005-3324-7290

mathgov.org

Affiliation

British University Vietnam (BUV)

Abstract

Contemporary governance frameworks and artificial intelligence alignment approaches repeatedly fail in three coupled ways: (i) they collapse plural values into single metrics that permit trading away fundamental rights, (ii) they underweight catastrophic tail risks through expected-value reasoning, and (iii) they remain vulnerable to specification gaming, where optimization targets proxies while degrading intended outcomes. This paper introduces MathGov, a universal ethical operating system grounded in Union-Based Reality (UBR), the stance that interconnection and nested unions, not isolated agents, provide the correct structural grammar for describing physical, biological, ecological, cognitive, and social systems. From this ontology, Union-Based Ethics (UBE) is operationalized as a five-level decision cascade applied to a 49-cell welfare matrix (seven unions by seven welfare dimensions). The cascade is: (1) a Non-Compensatory Rights Constraint (NCRC) that excludes rights-violating options except under explicitly declared emergency procedures, (2) a Tail-Risk Constraint (TRC) that excludes options with unacceptable catastrophic exposure using Conditional Value-at-Risk (CVaR_α), (3) a Containment Check that prevents local optimization from degrading the coherence and viability of containing unions, (4) a Ripple Logic Score (RLS) that ranks remaining options by weighted welfare impacts after ripple propagation, and (5) a structural tie-break and drift monitor using the Union Coherence Index (UCI) and Hollowing-Out Index (HOI) when RLS differences are within an uncertainty band. The framework further specifies a Sentience Gradient Protocol (SGP) with a rights plateau for managing intelligences, a Hybrid Democratic Weighting (HDW) scheme combining constitutional floors with democratic tuning, explicit uncertainty handling via sparse ripple kernels, and an auditable Provenance and Compliance Certificate (PCC) embedded in a Notice-Choose-Act-Reflect (NCAR) learning loop. We present MathGov as an implementable, corrigible decision system for individuals, organizations, governments, and AI systems, and outline a validation program with explicit falsification criteria.

Keywords: AI alignment, lexicographic ethics, existential risk, relational ontology, machine moral status, multi-scale governance

---

1. Introduction: The Need for a Universal Ethical Operating System

1.1 Alignment Failures Across Scales

Contemporary societies operate within a tightly coupled, high-dimensional environment where climate dynamics, global supply chains, digital communication networks, financial systems, and emerging artificial intelligence systems interact in ways that increasingly resist prediction or governance. Decisions taken at one organizational scale, such as corporate investment choices, national energy policy, or algorithmic deployment, propagate rapidly through multiple layers of human and ecological organization, generating consequences that conventional decision frameworks fail to anticipate or manage (Meadows, 2008; Steffen et al., 2015).

In this context, alignment is not exclusively an artificial intelligence problem. It represents a general challenge of ensuring that the actions of individuals, institutions, governments, and machine systems remain consistent with the protection of fundamental rights, the avoidance of catastrophic failure modes, and the long-term flourishing of sentient beings and the planetary systems that sustain them (Russell, 2019; Ord, 2020). The misalignment observable in AI systems, where optimization for narrow objectives produces unintended harms, reflects structural features present across governance domains: climate policies that prioritize short-term economic growth over biosphere integrity, corporate metrics that incentivize quarterly profits over worker well-being, and institutional designs that systematically externalize costs onto future generations or vulnerable populations (Raworth, 2017; Rockström et al., 2009).

Existing decision frameworks exhibit three recurring failure modes that together constitute what we term the alignment trilemma:

Scalarization of value. Many approaches collapse a rich ethical landscape into a single number, such as net monetary benefit, expected utility, or a composite index. Arrow's (1963) impossibility theorem demonstrates that no preference-aggregation rule can simultaneously satisfy minimal fairness criteria when preferences conflict fundamentally across multiple dimensions and scales. When rights and risks are folded into this scalar, decision-makers can trade off severe harms to some unions, such as marginalized communities or ecosystems, for gains elsewhere, without explicit, non-negotiable safeguards. Cost-benefit analysis, despite its utility in bounded contexts, systematically distorts ethical priorities by collapsing incommensurable values into monetary equivalents (Sen, 2009).

Tail-risk blindness. Standard expected-value reasoning tends to underweight extreme but low-probability harms, especially when probabilities and consequences are highly uncertain (Taleb, 2012). A 0.1 percent probability of civilization-ending catastrophe becomes algebraically conflated with a 0.1 percent improvement in quarterly output, a symmetry that is both morally untenable and empirically dangerous. Climate tipping points, pandemic risks, nuclear escalation, and AI misalignment exemplify threats that conventional frameworks systematically underweight (Intergovernmental Panel on Climate Change [IPCC], 2023; Bostrom, 2014). Coherent risk measures such as Conditional Value-at-Risk provide mathematically rigorous alternatives that respect the asymmetric nature of catastrophic outcomes (Artzner et al., 1999; Rockafellar & Uryasev, 2000).

Specification gaming and Goodhart vulnerability. When systems optimize proxy metrics under intense pressure, they exploit loopholes rather than achieving intended outcomes, a phenomenon formalized as Goodhart's Law (Goodhart, 1984; Manheim & Garrabrant, 2018). AI systems trained on narrow reward functions produce unintended harms (Amodei et al., 2016), while institutional metrics such as GDP growth incentivize ecological destruction (Rockström et al., 2023). Traditional remedies, including adding more constraints, refining metrics, or increasing oversight, often shift the gaming vector to new exploits rather than eliminating it.

These failures are not isolated bugs but symptoms of a deeper conceptual error: treating ethics as subjective preference aggregation rather than as the formal study of coherence conditions for nested living systems.

1.2 Requirements for a Universal Ethical Operating System

To address these structural failures, a universal ethical operating system must satisfy at least five requirements:

Rights-first, non-compensable constraints. Certain harms, especially those involving core rights to life, bodily integrity, and basic subsistence, must be treated as non-compensable. No amount of aggregate benefit elsewhere should mathematically justify their routine violation. This requires a formal structure that treats rights as lexicographic constraints, not as terms in a single weighted sum. Rawls (1971) established the priority of liberty over other social goods; MathGov extends this logic to encompass a broader set of rights across multiple organizational scales.

Explicit catastrophic risk bounding. The system must explicitly represent and bound tail risks to humanity and the biosphere. Rather than relying solely on expected values, it should employ risk measures that prioritize the avoidance of irreversible or existential outcomes. Ord (2020) estimates existential risk from various sources over the coming century; a decision framework that cannot represent such risks cannot adequately manage them.

Multi-dimensional, multi-scale welfare representation. Welfare cannot be reduced to a single dimension such as income or a single scale such as the nation-state. A valid framework must represent multiple welfare dimensions, including material, health, social, epistemic, agency, meaning, and environmental, and track them across nested unions from individual to biosphere. The capability approach developed by Sen (1999) and Nussbaum (2011) establishes the theoretical foundation for multi-dimensional welfare assessment; MathGov provides computational architecture for its implementation.

Computability and auditability. The framework must be implementable in software, with clearly defined inputs, intermediate computations, and outputs. It must generate an auditable record of decision rationale, allowing packaged review and enabling both human and machine agents to participate under the same rules. As AI systems increasingly participate in consequential decisions, shared ethical architecture becomes essential for coordination and accountability (Christian, 2020).

Recent AI governance and risk-management regimes emphasize documented risk management, transparency, and accountability obligations for AI systems. The National Institute of Standards and

Technology (NIST) AI Risk Management Framework (NIST, 2023), the Organisation for Economic Co-operation and Development (OECD) Recommendation on AI (OECD, 2019), and International Organization for Standardization/International Electrotechnical Commission (ISO/IEC) guidance on AI risk management (ISO/IEC, 2023) exemplify this trend. MathGov is designed to be interoperable with these regimes through its auditable PCC and its explicit treatment of risk, while adding two elements they typically do not formalize as decision operators: (i) a lexicographic, non-compensatory rights constraint (NCRC) and (ii) explicit catastrophic tail-risk bounding (TRC) using coherent tail-risk measures where applicable.

Corrigibility and learning. Given the limits of human knowledge and the complexity of real systems, any ethical operating system must treat itself as fallible and updateable. It should include explicit feedback loops for learning from outcomes, revising system parameters, and correcting mis-specifications in light of new evidence. This aligns with the scientific method's commitment to falsifiability and revision (Popper, 1959).

1.2.1 Implementation scope and external dependencies (Normative)

This paper provides the complete conceptual, formal, governance, and audit specification for all tiers. Tier 1 to 2 decisions can be executed using only in-paper defaults, notably Appendices AD, S, T, and AF, together with a valid PCC and AIL compliance. For docs-only execution at Tier ≤ 2, the appendices include embedded starter artifacts (e.g., Starter KOPS in Appendix S, starter rights anchors in Appendix T, and the DSL-20-TRAINING-V1 scenario library in Appendix D.7). Tier ≥ 3 and Tier-4 Pilot-Executable claims remain bound to governed registries referenced by hash per AIL.

Tier 3 to 4 execution additionally requires an external, hash-bound ProofPack bundle referenced in the PCC. For Tier-4 claims, the run MUST reference ProofPack registries and manifests by SHA-256 and MUST obey the ProofPack canonicalization profile(s) (including NO_FLOATS and exact rationals where required). Any illustrative values printed in this paper or the Appendices (including decimals) are non-canonical and MUST NOT be used for Tier-4 hashing.

ProofPack contents (rev14). The ProofPack provides: canonicalization profiles and hashing rules; JSON Schemas for PCC and registries; registry manifests and per-artifact hashes; and the canonical, machine-readable registries (e.g., rights anchors, catastrophe indicators, kernel edges, scenario library, and HDW ballots/weights where applicable) referenced by hash in PCCs. It does not provide executable replay tooling or a conformance harness in this manifest-only release.

Manifest-only definition (Normative). In rev14.1, ProofPack is manifest-only in the sense that it ships no executables. It DOES ship the canonical machine-readable JSON registries, schemas, manifests, and test vectors required for Tier-4 replay.

Rev14.10 Tier-4 binding decisions (Normative).
For rev14.1 and later, Tier-4 Pilot-Executable determinism is defined by the following binding decisions:
(D1) ProofPack manifest-only means “no executables shipped,” not “no data shipped.” The ProofPack bundle MUST ship hash-bound JSON data artifacts including registries, schemas, manifests, and test vectors (if any). Tooling MAY exist separately as a Pilot Kit, but tools are not part

of the ProofPack release.
(D2) Canonical cell identifiers are object-typed: { "u": <int 1..7>, "d": <int 1..7> }. Human-readable cell names are display-only and MUST NOT be used as identifiers in Tier-4 registries or PCC snapshots.
(D3) Array ordering: arrays are hashed exactly as stored. Canonicalization does not reorder arrays. Registries MUST be authored in canonical order; tooling MAY validate order but MUST NOT rewrite arrays.
(D4) Registry hash anchoring: registry files are hashed as canonical JSON bytes. Hash values MUST be recorded externally in the ProofPack MANIFEST indexes. Registry files MUST NOT contain self-referential hash fields.

Tier-4 evaluation scope. Claims of Tier-4 Pilot-Executable compliance MUST be evaluated against this Foundation Paper plus the Appendices plus ProofPack v1.0 at the referenced revision. The Foundation Paper remains the single source of truth for decision logic and governance requirements. The ProofPack provides the artifact specifications required for deterministic replay.

1.3 MathGov and Union-Based Reality in Context

MathGov is built on UBR, which models existence as a network of interdependent unions rather than isolated individuals or agents. UBR does not replace systems theory, network science, or relational ontologies; it operationalizes their shared implication of multi-scale interdependence into a computable and auditable decision architecture. These unions range from the individual self to global humanity and the biosphere, with meta-unions (Cosmic and Universal) available for deep-time reasoning. On top of this ontology, MathGov defines a structured welfare space and a lexicographic decision cascade operating over a 49-cell matrix: seven operational unions crossed with seven welfare dimensions.

Formally, for each candidate action a, MathGov evaluates impacts over the welfare space indexed by union u and dimension d. Each action induces (i) a direct impact field and (ii) a propagated impact field, where propagation is computed using a sparse ripple kernel $\mathbf{K}$ that encodes cross-effects between cells in the union-dimension matrix. Candidate actions are filtered in lexicographic order: first by the NCRC, then by TRC using a CVaR-based bound on catastrophic exposure, then by a Containment Check that ensures local gains do not degrade containing unions, and only then aggregated into an RLS. When top options are statistically indistinguishable in RLS, a UCI tie-break evaluates structural coherence effects to support final selection.

MathGov includes an SGP with a rights plateau for Managing Intelligences (MI). Once a system, whether biological, digital, or hybrid, crosses a specified threshold of sentience and governance capability, it is assigned full rights protection. Greater intelligence above that plateau does not confer additional rights, only additional capabilities and responsibilities.

Value aggregation is handled via HDW. HDW combines constitutionally protected floors for unions and welfare dimensions with democratic tuning for the remaining weight mass. This strikes a balance between technocratic safety and democratic legitimacy, addressing the tension between expert knowledge and popular sovereignty that characterizes contemporary governance debates (Dryzek, 2000).

All non-trivial decisions yield a structured PCC, enabling audit, red-teaming, and long-term learning. The system operates within a NCAR loop that makes MathGov a living, corrigible framework rather than a static doctrine.

MathGov can be positioned relative to existing ethical frameworks and AI alignment approaches. The following table summarizes how MathGov addresses each component of the alignment trilemma compared to existing approaches:

Failure Mode	Existing Frameworks	MathGov Solution
Value Pluralism Intractability	Collapse into single metrics (utilitarianism, CBA)	Lexicographic cascade: NCRC → TRC → Containment → RLS → UCI (no compensation across levels)
Tail-Risk Neglect	Expected utility, cost-benefit analysis	TRC with CVaR_α: Explicit catastrophic risk bounding
Specification Gaming	Proxy metrics (GDP, AI rewards)	Multi-layer constraints + Containment + UCI/HOI monitoring

Unlike classical utilitarianism, MathGov does not maximize aggregate welfare without threshold protections; it imposes non-compensatory floors via NCRC, treating certain harms as inadmissible regardless of offsetting benefits. Unlike pure deontology, MathGov pays explicit attention to consequences and cross-scale dynamics, providing a complete decision procedure through the lexicographic cascade. Unlike Rawlsian justice, which focuses on basic societal structure at the nation-state level, MathGov addresses all union scales simultaneously. Unlike Constitutional AI approaches (Bai et al., 2022a), which train systems on principles without formal specification, MathGov provides mathematically explicit constraints and optimization criteria. Unlike Reinforcement Learning from Human Feedback (RLHF) approaches (Bai et al., 2022b), which rely on human feedback to shape behavior, MathGov provides structural constraints that cannot be optimized away.

MathGov shares structural similarities with Multi-Criteria Decision Analysis (MCDA) frameworks, particularly those employing lexicographic extensions (Belton & Stewart, 2002; Keeney & Raiffa, 1976). However, MathGov differs from standard MCDA in several respects: (i) it integrates explicit tail-risk constraints using coherent risk measures rather than treating risk as one criterion among many, (ii) it embeds a containment principle that prevents sub-system optimization from degrading containing systems, (iii) it provides explicit protocols for handling rights as non-compensable constraints rather than as weighted criteria, and (iv) it includes a learning loop (NCAR) and audit architecture (PCC) designed for corrigibility and institutional accountability rather than one-shot decision support.

1.3.1 Addressing the Relocation Objection (Where Value Judgments Live)

A natural critique is that MathGov does not eliminate value disagreement, it relocates it from explicit preference aggregation into framework design choices: the union set, the welfare dimensions, rights thresholds, catastrophe definitions, masking rules, and weight floors. MathGov accepts this in one sense: no non-trivial governance system can avoid normative commitments. The relevant question is not whether commitments exist, but whether they are (i) explicit rather than hidden, (ii) auditable and versioned rather than ad hoc, (iii) protected against capture, and (iv) corrigible under evidence without rewriting history.

MathGov therefore does not claim to discover a single "true" utility function. Instead, it constitutionalizes certain protections (NCRC and TRC) as lexicographic admissibility constraints, and it restricts democratic value aggregation (HDW) to the remaining degrees of freedom inside an explicit constitutional envelope (floors, non-maskable cells, and anti-capture rules). This changes the location and tractability of normative choice: contestation moves from opaque scalarization choices (for example, discount rates and proxy metrics) into explicit, governed parameters that can be debated, tested, audited, and revised through NCAR and Charter procedures.

In short, MathGov does not remove ethical disagreement. It makes the structure of disagreement computationally explicit, non-compensatory where necessary (rights and catastrophic exposure), and corrigible through transparent governance rather than hidden scalarization.

1.4 Structure of the Paper

The remainder of this paper is organized as follows. Section 2 introduces the ontological foundation of UBR, defining unions, their nesting, and the role of meta-unions. Section 3 presents the normative foundation, including the Minimal Normative Axiom (MNA), operational definitions of help and harm, and the Containment Principle. Section 4 provides an overview of the MathGov architecture, including the welfare matrix, lexicographic cascade, and NCAR loop. Sections 5 through 8 formalize the welfare space, rights floors (NCRC), TRC, and ripple propagation via the kernel. Section 9 defines the SGP and the MI rights plateau. Section 10 details the HDW scheme. Section 11 presents the scoring and selection procedure under uncertainty, including RLS, UCI, and HOI. Section 12 describes the NCAR learning loop. Section 13 introduces the PCC and audit layer. Section 14 discusses validation and falsification. Sections 15 and 16 address applications and limitations. Section 17 concludes with implications for multi-scale alignment.

Methodological approach and AI assistance. This paper follows a design-science and normative-engineering methodology. It specifies an implementable decision architecture, defines its mathematical objects and constraints, analyzes failure modes (including rights violations, tail-risk underweighting, and metric gaming), and proposes auditable artifacts (the PCC) and validation criteria for empirical testing in pilot deployments. Generative AI tools (OpenAI ChatGPT and Anthropic Claude) were used as writing and reasoning assistants during drafting, revision, and consistency checking. The author reviewed, verified, and edited all outputs and assumes full responsibility for the content, claims, and citations.

2. Ontological Foundations: Union-Based Reality

2.1 Reality as Relational, Not Atomic

The foundational premise of MathGov is an ontological claim about how entities exist in the world:

Ontological Thesis. No entity exists in complete isolation. Every entity is embedded in networks of interaction that partially constitute its identity, constrain its behavior, and transmit the consequences of its actions.

This thesis is descriptive rather than normative. It synthesizes converging evidence from multiple empirical domains, as illustrated in the framework overview.

The Framework Overview: Convergent Evidence for Union-Based Reality

Across disciplines, including biology, ecology, cognitive science, and systems theory, empirical evidence supports the UBR thesis: reality is fundamentally relational, with entities embedded in nested networks of interaction. This convergence justifies UBR as the ontological foundation for MathGov.

Note on cross-domain convergence and scope conditions. UBR is adopted here as a modeling stance: in high-coupling systems, outcomes are shaped by interaction structure and cascading effects, and models that treat agents as isolated optimizers systematically misrepresent consequences. The empirical grounding most directly relevant to governance comes from systems science, network science, institutional dynamics, ecology, and cognitive/social interdependence research. References to relational structure in physics are best understood as illustrative analogy, not as a derivation of governance claims.

UBR is most applicable in domains characterized by: (i) high coupling between entities, where actions propagate through networks; (ii) significant externalities, where consequences extend beyond the acting agent; (iii) multi-scale feedbacks, where micro-level actions aggregate to macro-level effects; and (iv) long time horizons, where delayed consequences matter. UBR does not claim that all systems require relational modeling; low-coupling, short-horizon, single-agent decisions may be adequately handled by simpler frameworks.

In biology, living organisms emerge through symbiotic partnerships spanning cellular to ecosystem scales. The endosymbiotic origins of complex cells reveal that mitochondria originated as independent bacteria that entered into cooperative relationships with ancestral cells approximately 1.5 billion years ago (Margulis, 1998; Gray, 2012). The holobiont concept extends this further: humans are human-microbial ecosystems, harboring vastly more bacterial genes than human genes (Zilber-Rosenberg & Rosenberg, 2008). Trillions of bacteria in the gut influence cognition, emotion, and behavior (Cryan & Dinan, 2012).

In ecology, the reintroduction of wolves to Yellowstone triggered cascading effects across the entire ecosystem (Estes et al., 2011). Biogeochemical cycles demonstrate that carbon, nitrogen, phosphorus, and water cycle through living organisms, atmosphere, oceans, and lithosphere in integrated planetary metabolism. Planetary boundaries research identifies nine boundaries within which humanity can safely operate, demonstrating that transgressing boundaries in one domain cascades to affect others (Rockström et al., 2009, 2023).

In cognitive science, human neocortex size correlates with typical group size across primates, suggesting that human cognitive evolution was driven by social complexity (Dunbar, 1993). Attachment theory demonstrates that secure attachment relationships constitute the necessary substrate for healthy psychological development (Bowlby, 1988). Social contagion research shows that emotions, behaviors, and beliefs spread through social networks with predictable patterns (Christakis & Fowler, 2009).

In systems science, complex systems organize into hierarchies where interactions within modules prove stronger than interactions between modules, enabling both local adaptation and global coordination (Simon, 1962). Network theory reveals that real-world networks exhibit universal structural features across physical, biological, and social domains (Newman, 2010; Barabási & Albert, 1999).

If this thesis is correct, then ethical or governance frameworks that treat agents as isolated optimizers will systematically misrepresent both the consequences of actions and the conditions for long-term stability. Decisions taken at one node, such as a person, a firm, or an AI system, propagate through networks, affecting other nodes in sometimes unforeseen ways. An ethical operating system that ignores this structure is incomplete.

2.2 Unions as the Unit of Analysis

To make relational reality tractable, MathGov introduces the concept of a union.

Definition. A union is a bounded pattern of interdependence: a set of entities whose interactions with each other are significantly stronger, more frequent, or more consequential than their interactions with entities outside the set.

Unions are not metaphysical substances. They are analytical constructs designed to carve reality in ways that capture major patterns of causal influence and shared welfare. The choice reflects a design decision based on structural sufficiency for viability analysis rather than a claim of mathematical necessity.

MathGov uses seven operational unions as primary layers, indexed by $u \in \{1, 2, 3, 4, 5, 6, 7\}$:

Self (U_1). The individual conscious agent, biological or digital, as the locus of subjective experience and decision-making. Empirical grounding comes from consciousness studies (Tononi, 2008; Dehaene, 2014) and unified agency research. Characteristic timescale ranges from seconds to decades.

Household (U_2). The primary unit of cohabitation and resource pooling, including families, shared living arrangements, or other small, tightly coupled domestic units. Empirical grounding comes from Dunbar's intimate group research (Dunbar, 1992) and household economics (Becker, 1981). Characteristic timescale ranges from days to decades.

Community (U_3). Local networks of direct, repeated interaction: neighborhoods, villages, schools, professional networks with strong mutual influence. Empirical grounding comes from social capital research (Putnam, 2000) and the social brain hypothesis (Dunbar, 1993). Characteristic timescale ranges from months to generations.

Organization (U_4). Purpose-driven collectives with formal structure: firms, NGOs, universities, government agencies. Empirical grounding comes from organizational behavior (March & Simon, 1958) and institutional economics (North, 1990). Characteristic timescale ranges from years to centuries.

Polity (U_5). Political units with legitimate authority over a jurisdiction: cities, states, nations, or equivalent governance entities. Instances include municipalities, provinces, nations, regional blocs (EU, ASEAN, AU), and global intergovernmental bodies (UN, WHO, WTO). Empirical grounding comes from political science and state theory (Weber, 1978). Characteristic timescale ranges from decades to centuries.

Humanity/CMIU (U_6). The Collective Managing Intelligence Union (CMIU) encompasses current humanity and all sufficiently advanced intelligences sharing capacity for moral reasoning and coordinated governance. CMIU sits above and provides coordination context for all Polity instances; it represents humanity's collective capacity to address global challenges. Empirical grounding comes from global systems research and international relations theory. Characteristic timescale ranges from generations to millennia.

Biosphere (U_7). Earth's integrated living systems including atmosphere, hydrosphere, and lithosphere as they interact with and support life. Empirical grounding comes from Earth system science (Steffen et al., 2015) and ecology (Odum, 1971). Characteristic timescale ranges from centuries to geological epochs.

These seven operational unions balance completeness against tractability. Fewer levels would miss critical scales of decision and impact; more levels would increase computational complexity without proportionate gains in clarity. The 49-cell structure that emerges when these unions are crossed with seven welfare dimensions provides the foundational matrix of MathGov.

A crucial clarification: the seven operational unions are structural types, not monolithic entities. Each union type comprises many instances. Household includes billions of distinct household instances worldwide. Organization includes millions of distinct organizations. Polity includes instances at multiple geographic scales. Analyses may disaggregate impacts across relevant instances for diagnostic purposes, but for the 49-cell matrix evaluation, impacts are aggregated within the appropriate union type row.

2.2.1 Union Types Versus Instances: Aggregation Rules

Implementation note (instances). An “instance” is a concrete, countable member of a union type at a declared scope (who/where/when). Examples: a specific household, a particular organization, a municipality, or a defined community catchment. Each PCC MUST declare the instance set used (instance_id, scope, population n_i, and membership rules for multi-parent cases), so independent auditors can reproduce instance aggregation and worst-off subgroup checks.

Pilot guidance. For early Tier-4 pilots, you may set instances(u) to a single scoped instance for each union type (e.g., the org running the pilot, its local community catchment, the relevant polity), provided the PCC explicitly documents the boundary choice and any excluded populations as a limitation. For broader deployments, use the population-weighted aggregation rule as written.

The seven operational unions are structural types, each comprising many instances. When evaluating impacts, the following aggregation rules apply:

Default Aggregation Method: Population-weighted mean

Let I_i denote the impact for instance i within union type u. The aggregated impact for the union type is:

I_bar_u = (sum_{i in instances(u)} n_i * I_i) / N_u

N_u = sum_{i in instances(u)} n_i

where n_i is the population of instance i and N_u is the total population of union type u.

Alternative Methods (require PCC justification):

Method	Formula	When Appropriate
Worst-off instance	min over instances	Rights-adjacent analysis
Stakeholder-weighted	Expert-assigned weights	Complex multi-stakeholder
Equal-weighted	Simple mean	Instances of similar scale

Rights Exception: For rights-covered cells, always use worst-off subgroup analysis within instances, then aggregate using the specified method.

2.3 Meta-Unions for Long-Horizon Reasoning

Beyond the operational seven, MathGov recognizes two meta-unions:

Cosmic Union (U_8). The broader physical environment beyond Earth: the solar system, near space, and eventually interstellar contexts. At present, most human decisions have negligible direct impact at this scale, so MathGov does not parameterize Cosmic Union in standard calculations. As human and digital civilizations expand beyond Earth through space infrastructure and off-world settlements, the Cosmic Union can be brought into the formalism as an eighth operational union with its own welfare indicators.

All-Encompassing Infinite Union (AIU, U∞). The conceptual union consisting of all existence: every physical, informational, and possibly trans-physical entity. AIU functions as a philosophical boundary condition rather than a computational object. It reminds us that any local model is embedded in a larger reality that we cannot fully parameterize or empirically access. MathGov treats AIU as not yet testable; it informs humility, not direct computation.

These meta-unions do not participate in routine scoring but are relevant when considering deep-time, cosmological, or metaphysical considerations, for example in discussions of long-term trajectories of intelligence in the universe or scenarios with cosmic stakes.

The complete union taxonomy is therefore:

- Operational: U_1 through U_7
- Meta: U_8 (Cosmic), U_∞ (AIU)

Default computations in MathGov use U_1 through U_7 unless a PCC explicitly enables a parameterized Cosmic union or invokes Universal for philosophical boundary reasoning.

Meta-Union Extension Protocol. When U_8 (Cosmic) is activated as an operational union, the following modifications apply:

(a) Matrix expansion: The welfare matrix expands from 49 cells (7 × 7) to 56 cells (8 × 7). The kernel **K** expands from 49 × 49 to 56 × 56, with new entries initialized to zero unless explicitly specified.

(b) Rights extension: The canonical rights set extends to include Cosmic cells only for rights where the extension is semantically meaningful. ECOL (Ecological Integrity) extends to include U_8 for space environment protection. Other rights extend only with explicit governance justification.

Serialization note (Tier-4): whenever a PCC or registry serializes a welfare cell, it MUST use { "u": <int 1..7>, "d": <int 1..7> }.

(c) Catastrophe cell set: The base catastrophe set C_cat may be extended to include (U_8, D_7) (Cosmic-Environment) when decisions have plausible space-scale consequences.

(d) Governance procedure: Activation of U_8 as an operational union requires Charter-level approval with documented justification and specification of which cells, rights, and catastrophe considerations apply.

2.4 Nested Membership and Non-Separability

In practice, unions are nested. A given Self belongs to a particular Household, which is embedded within one or more Communities, Organizations, and Polities, all nested within Humanity and the Biosphere. The welfare of each union is not independent of its containing unions.

Let W_u denote the welfare of union *u*. Then, for the Self:

W_Self = f(W^intrinsic_Self, W_Household, W_Community, W_Organization, W_Polity, W_Humanity/CMIU, W_Biosphere).

where W^intrinsic_Self represents factors directly tied to the individual, and the remaining terms represent the welfare of containing unions. A person's long-term welfare depends on their household, local community, organizational context, polity, species-level systems, and biosphere. No amount of individual wealth can fully compensate for collapse of the biosphere or breakdown of basic social cohesion.

Similarly, the welfare of higher-level unions partly depends on the state and behavior of lower-level ones. A polity with citizens in chronic ill-health or epistemic fragmentation will struggle to maintain resilience and coherence; a biosphere under severe anthropogenic stress ultimately threatens humanity and all nested unions.

This nested structure is the basis for MathGov's insistence on evaluating ripple effects: actions at one level often have non-trivial consequences at multiple other levels, which must be represented explicitly.

Multi-Parent Union Membership. Real institutional membership is not a single chain; individuals may belong to multiple communities, organizations, and even polities (through dual citizenship or multi-jurisdictional residence). MathGov handles this through the following protocol:

(a) Default chain: For simplicity, the canonical nesting chain $U_1 \subset U_2 \subset U_3 \subset U_4 \subset U_5 \subset U_6 \subset U_7$ is used as the default for containment checks and ancestor functions.

(b) Instance-level analysis: When a decision materially affects multiple instances of the same union type (e.g., multiple communities), the impact analysis should disaggregate impacts across relevant instances before aggregating to the union type row.

(c) Ancestor function: The function Anc(u, D) returns containing unions up to depth D using the default chain. For decisions where multi-parent membership is material, the PCC should document which containing instances are considered and how conflicts are resolved.

(d) Worst-off subgroup protection: For rights checks, disaggregation across affected subgroups is required (see Section 3.2.8), ensuring that multi-parent complexity does not allow rights violations to be averaged away.

2.5 Constructive and Pathological Unions

Because unions are defined by patterns of interdependence, they can be constructive or pathological. Constructive unions, such as healthy communities and trustworthy organizations, contribute positively to the welfare of their members and the unions that contain them. Pathological unions, such as organized crime networks, malignant tumors, or hostile disinformation ecosystems, may enhance welfare for some internal members in narrow dimensions while degrading welfare at higher levels.

MathGov does not assume that any union is good by definition. Instead, union quality is evaluated through its ripple effects, subject to constraints on rights and catastrophic risk. The Containment Principle (formalized in Section 3.4) makes this explicit: improvements in a sub-union's welfare are not automatically counted as good if they damage the coherence or viability of containing unions. This addresses cases like cancer, corruption, or extractive industries that profit a subset while undermining foundational systems.

3. Normative Foundations: The Minimal Normative Axiom and Its Operationalization

3.1 The Minimal Normative Axiom

MathGov makes exactly one explicit normative commitment:

Minimal Normative Axiom (MNA). Sentient flourishing matters. Unnecessary suffering should be reduced. The conditions that enable continued flourishing should be preserved.

This axiom is not derived from physics, biology, or any descriptive claim about the world. It is a normative stance, a declaration of what we take to be ethically significant. The MNA is minimal in three precise senses:

Content-minimal. It makes no specific claims about what flourishing consists of beyond the continuation of sentient existence with conditions for well-being. It does not prescribe a culturally specific or teleological conception of the good life.

Scope-bounded. It applies only to agents and institutions that accept that sentient experience matters. Agents that reject the MNA are outside the framework's normative scope. MathGov does not claim to refute such positions; it simply does not attempt to govern them.

Derivationally sufficient. Given empirical facts about union structure (Section 2), the MNA provides sufficient foundation for deriving the substantive constraints and procedures that follow.

3.1.1 The Conditional Is-Ought Bridge

The transition from descriptive ontology (UBR) to a normative framework (UBE) requires explicit philosophical care to avoid the naturalistic fallacy, identified by Hume (1739/1978) as the invalid derivation of normative claims directly from descriptive facts. MathGov addresses this problem by introducing a conditional is-ought bridge: normative obligations are not inferred from the structure of reality alone, but arise only once a minimal normative commitment is explicitly adopted.

If (i) reality is union-structured such that actions propagate through nested unions (UBR), and (ii) sentient suffering and flourishing are morally significant (MNA), then agents and institutions ought to evaluate choices by their cross-union impacts and by whether they preserve the enabling conditions for continued flourishing.

The normativity enters only through the MNA. UBR specifies the structural pathways along which that axiom applies. It tells us where consequences flow, not whether consequences matter. This separation grounds MathGov's logical coherence and prevents descriptive claims about interdependence from being mistaken for ethical conclusions.

3.1.2 Reader Clarifications (Expository, Non-Formal)

Q: Is the MNA a claim about what flourishing consists of? A: No. The MNA is a procedural commitment: if sentient flourishing matters, then we should act to preserve the conditions for it. It constrains admissibility and decision procedure; it does not prescribe specific content.

Q: Does the MNA assume all cultures share this commitment? A: No. The MNA is conditional: if an agent or culture accepts that sentient suffering and flourishing matter, then MathGov provides a coherent way to operationalize that commitment. Cultures that reject this premise place themselves outside MathGov's scope.

Q: How is the MNA different from utilitarianism? A: Utilitarianism maximizes aggregate welfare without structural constraints. The MNA prioritizes rights and catastrophic risk avoidance before welfare optimization (via the lexicographic cascade), and it explicitly models multi-scale interdependence rather than collapsing welfare into a single scalar.

3.2 Operational Definitions: Impacts, Admissibility, and Comparative Ranking

To implement the MNA in a way that is computable, auditable, and non-ambiguous, MathGov separates three distinct logical layers:

- Descriptive impact claims (what an option does to welfare),
- Deontic status (whether the option is permitted or forbidden), and
- Comparative choice (which option is better or worse among permitted alternatives).

This separation is not merely stylistic. It prevents the logical error of conflating what is the case with what ought to be done, and it enables the lexicographic cascade (Section 4.2) to function correctly: admissibility is determined before comparative ranking, and no amount of comparative advantage can override a failure of admissibility.

3.2.1 Impact Objects (Descriptive, Pre-Normative)

Let *a* be a candidate option. For each union *u* and dimension *d*, MathGov represents the post-propagation, post-saturation welfare impact as:

$$I_{u,d}(a) \text{ in } [-1, +1] \text{ for all } (u,d) \text{ in } U \text{ x } D.$$

Let $I(a) = [\ I_{u,d}(a)\ ]$ be the 7x7 post-saturation impact matrix for option a.

where:

- 0 means no change from baseline,
- Positive values indicate improvement in cell (u, d),
- Negative values indicate degradation in cell (u, d),

all under the calibration protocol defined in Section 5 and the ripple propagation mechanism of Section 8.

Scenario-conditioned impacts. When scenario evaluation is enabled (Section 7), impacts are computed per scenario:

I^s_{u,d}(a) in [-1,+1] denotes the impact in cell (u,d) under scenario s.

and the scenario-expected impact used for scoring is:

I_{u,d}(a) = sum_{s in S} p_s * I^s_{u,d}(a)

where p_s ≥ 0 and Σ_s p_s = 1.

Temporal scope. Temporal scope is captured through instance time horizons t_k and the temporal weighting function $\tau(t)$ (Section 5.2), with scenario modeling providing additional long-horizon structure (Section 7).

Sentience placement. Sentience does not appear as a union-level multiplier in RLS. It enters upstream, within-cell, when forming cell impacts from underlying entities and indicators (Sections 9.4-9.5), and only then is propagated and saturated into the final impact value.

3.2.2 Help and Harm (Sign-Consistent Magnitudes)

MathGov defines help and harm as non-negative quantities derived from impacts:

Help_{u,d}(a) = (I_{u,d}(a))^{+} = max(I_{u,d}(a), 0)

Harm_{u,d}(a) = (-I_{u,d}(a))^{+} = max(-I_{u,d}(a), 0)

where $(x)^+$ = max(x, 0).

Interpretation:

- Help_{u,d}(a) is the magnitude of improvement in cell (u, d).
- Harm_{u,d}(a) is the magnitude of degradation in cell (u, d).

These are purely descriptive quantities. They do not, by themselves, determine permissibility or comparative ranking. This decomposition ensures that "help" and "harm" in natural language map cleanly to non-negative mathematical objects, eliminating sign confusion in subsequent aggregation.

3.2.3 Admissibility: Permitted vs. Forbidden (Deontic Status)

MathGov treats rights constraints and catastrophic tail-risk as admissibility filters, not as terms in a weighted sum. This is the formal expression of non-compensability: no welfare gain can override a rights violation or an unacceptable catastrophic risk.

Define the admissibility predicate:

Rule (clarity): NCRC and TRC determine permissibility (Admissible). Containment determines selectability (Selectable). RLS ranks selectable options.

$\text{Admissible}(a) := \text{NCRC}(a) \land \text{TRC}(a)$

$\text{A\_adm} := \{ a \in A : \text{Admissible}(a) \}$

where:

1. NCRC(a) is the Non-Compensatory Rights Constraint (Section 6), and

2. TRC(a) is the Tail-Risk Constraint (Section 7).

Then:

1. **Permitted (admissible):** Admissible(a) = true

2. **Forbidden (inadmissible):** Admissible(a) = false

Critical implications:

1. An option can contain many "helps" and still be forbidden (e.g., because it violates a rights floor or exceeds the catastrophe corridor).

2. An option can include some harms and still be permitted, provided it stays within rights floors and tail-risk bounds.

This is the formal expression of "no compensation across levels": later welfare optimization never overrides rights or catastrophic safety.

Empty admissible set. Define the following sets:

If $\text{A\_adm} = \emptyset$, MathGov does not silently choose the "least bad" forbidden option. Instead, it triggers governed exception-handling in which any selection from inadmissible options is explicitly declared as an emergency deviation, minimized by lexicographic criteria, and paired with mandatory remediation and review:

1. If $\text{A\_NCRC} = \emptyset$, invoke NCRC Emergency Mode (Section 6.4), selecting the option that lexicographically minimizes the rights-violation vector and requiring a remediation plan and review cadence.

REG-RIGHTS-PRIORITY-v1 (Normative).

The canonical rights priority ordering used for Emergency Mode and any other lexicographic rights resolution is an ordered registry object:

REG-RIGHTS-PRIORITY-v1 := [LIFE, BODY, ECOL, LBTY, NEED, DIGN, PROC, INFO].

Emergency Mode binding (Normative). When Emergency Mode is invoked, option evaluation and selection MUST follow REG-RIGHTS-PRIORITY-v1 in order, and MUST NOT rely on any implicit or ad hoc ordering.

2. If A_NCRC ≠ ∅ but A_adm = ∅ (i.e., all NCRC-passing options fail TRC), invoke TRC Fallback Mode (Section 7.5), selecting the option with minimal catastrophic-tail exposure (minimal CVaR_ɑ) together with mandatory mitigation and enhanced monitoring.

3. In all cases, the PCC must record the admissibility failure and the emergency justification.

This structure prevents gaming via deliberate constraint construction while preserving auditability and human oversight.

Explicit Emergency Decision Tree. When the admissible set is empty, MathGov applies the following resolution algorithm:

Case 1: A_NCRC = ∅ (All Options Violate Rights)

When no option respects all rights constraints, MathGov applies a strict lexicographic minimization procedure over rights violation depths:

Step 1: Construct the violation depth vector. For each option *a*, compute the violation depth for each right in priority order:

where v_r(a) is defined in Section 6.3.

Step 2: Lexicographic comparison. Compare options lexicographically on their violation depth vectors. Option *a* is preferred to option *b* if, at the first index *i* where they differ, v_i(a) < v_i(b).

Formally, a ≻_lex b iff there exists *i* such that:

1. for all j < i, v_j(a) = v_j(b), and
2. v_i(a) < v_i(b)

Step 3: Secondary criterion (CVaR tie-break). Among options tied on the violation depth vector (identical depths for all rights), minimize CVaR_α to ensure tail-risk protection even in emergency mode.

Step 4: Tertiary criterion (RLS tie-break). Among options still tied after CVaR, maximize RLS.

Documentation requirements: Emergency Mode PCC must include:

1. Declaration of crisis conditions and triggering event
2. Complete violation depth vector for each option
3. Lexicographic comparison showing selection rationale
4. Planned return-to-normal triggers and timeline

Note on "count" language: The lexicographic procedure compares violation depths (continuous magnitudes), not violation counts (binary indicators). An option with one severe violation (high depth) at a low-priority right may be preferred over an option with one moderate violation at a high-priority right. This respects the priority ordering while accounting for severity.

Case 2: A_NCRC ≠ ∅ but A_adm = ∅ (Rights-Respecting Options All Fail TRC)

When rights-respecting options exist but all exceed tail-risk threshold:

1. **Primary Criterion:** Among A_NCRC options, minimize CVaR_α. Select the rights-respecting option with lowest tail risk, even if above threshold.
2. **Secondary Criterion:** Among options tied on CVaR, maximize RLS.
3. **Mandatory Mitigation:** Selection of above-threshold option requires concurrent adoption of risk mitigation measures and explicit plan to return within threshold.
4. Escalation: TRC-emergency decisions require one-tier-higher approval (Tier 3 decision requires Tier 4 oversight).

Case 3: Both Failures Simultaneously

When all options violate rights AND all options exceed TRC threshold:

1. Apply Case 1 algorithm (rights-minimization primary). Rights protection takes absolute precedence even in cascading failures.

The selected option will also have the lowest feasible CVaR among options with equivalent rights-violation profiles, because Case 1 Step 3 uses CVaR as the secondary criterion.

2. Require highest-tier emergency oversight (organizational executive/board level).
3. Mandatory 24-hour reassessment for ongoing decisions.

3.2.4 Better vs. Worse (Comparative Ranking Among Admissible Options)

Among admissible options, MathGov produces a preference ordering using the RLS and tie-breakers (Sections 11.1-11.6). Let $\succ$ denote the induced preference relation ("*a* is preferred to *b*").

Canonical ranking rules:

1. If both *a* and *b* are admissible, then typically a ≻ b when RLS(a) > RLS(b), subject to uncertainty handling (δ-discrimination threshold, Judgment Calls) and integrity tie-breaks (UCI/HOI).

2. If *a* is admissible and *b* is not, then *a* strictly dominates *b* in the lexicographic cascade regardless of RLS.

3. If neither *a* nor *b* is admissible, neither enters the comparative ranking; see the empty admissible set protocol above.

Connection to Help/Harm: The RLS aggregates cell-level impacts (which decompose into Help and Harm components) weighted by union weights w_u, dimension weights v_d, and the applicability mask m_{u,d}. Weights w and v are governance inputs (HDW, Section 10) and do not alter admissibility; they only rank within A_adm. The ought enters only through the MNA. UBR identifies the structural pathways along which that axiom applies. It tells us where consequences flow, not whether consequences matter. This separation is the foundation of MathGov's logical coherence.

Propagation–Masking Canonical Rule (Normative).

All computations of direct impacts and propagated impacts MUST be performed in the full state space (49-cell welfare vector, or 56-cell if a Cosmic layer is enabled).

Applicability masks m_{u,d} MAY be used to exclude cells from RLS aggregation only. Masks MUST NOT be applied inside the propagation operator and MUST NOT be used to weaken or bypass NCRC or TRC checks.

Cell multipliers κ_{u,d} MAY be used as a declared per-cell scaling factor in RLS aggregation. κ_{u,d} defaults to 1.0. κ_{u,d} MUST NOT change which cells are active (m_{u,d}) and MUST NOT be used to bypass NCRC or TRC. If any κ_{u,d} ≠ 1.0 is used, the PCC MUST record the values and justification.

Audit requirement. If masking is used, the PCC MUST record: (i) the mask schema (cells masked/unmasked), (ii) the rationale, and (iii) confirmation that masked cells remain present in the run record and are reportable for review.

3.2.5 "Good" and "Bad" as Derived Labels (Not Primitives)

To preserve natural language intuitions without creating logical ambiguity, MathGov treats "good" and "bad" as derived procedural labels, not as primitive terms in the formal system:

1. **Forbidden-bad:** Any option with Admissible(a) = false.

2. **Permitted-better / Permitted-worse:** Comparative labels within the admissible set using the ≻ relation.

3. **Good (procedural):** An option selected by the MathGov selection rule from the admissible set that passes the containment check, under the declared uncertainty policy.

4. **Good-with-override (procedural):** An option selected despite a containment failure, with explicit escalation approval and documented justification recorded in the PCC. This label signals that the selection required governance intervention beyond the standard cascade.

5. **Bad (procedural):** Either forbidden-bad, or admissible-but-dominated by another admissible option.

This definitional strategy avoids conflating two mathematically distinct objects:

1. A predicate (admissible/forbidden), and

2. An ordering (better/worse among admissible options).

End-to-end execution algorithm (Implementation Pseudocode) (Normative).

Input: option set O, baseline x_0, indicator mappings + anchors, scenario set S with probabilities, weights, kernel K (or KOPS), applicability mask m (optional for RLS only), rights coverage sets C_r, rights thresholds θ_r, TRC parameters.

1) For each option a ∈ O: construct direct impacts using baseline-delta impacts (Baseline-Zero Rule).

2) Propagate impacts in full state space (Quick or Full) to obtain propagated impacts. (Tier 4 Pilot-Executable rev14.x: Quick only; see §13.8.2.)

3) Rights check (NCRC): evaluate worst-off subgroups where feasible; if infeasible at Tier-3, apply γ_subgroup conservative bound for rights checking only. If any right violates its threshold, a is inadmissible.

4) Tail-risk check (TRC): compute scenario losses, compute CVaR (or declared TRC mode), and test corridor admissibility. If TRC fails, a is inadmissible.

5) If no admissible options remain: invoke Emergency Mode and/or TRC fallback per the spec, governed by REG-RIGHTS-PRIORITY-v1.

6) Compute RLS for admissible options: apply applicability mask only at aggregation; compute discrimination band checks.

7) Apply containment gating (Mode A) to determine Selectable(a). Mode B is diagnostic-only.

8) Select argmax over Selectable(a) using declared tie-breaks (UCI/HOI) when applicable and available; otherwise escalate per UCI Unavailability Rule.

9) Emit PCC with all required snapshots, registries, audit flags, and conformance claims.

The distinction between "Good (procedural)" and "Good-with-override (procedural)" is essential for audit purposes. When a decision proceeds despite containment failure, the override must be traceable, justified, and subject to heightened scrutiny during NCAR reflection cycles. This prevents the erosion of containment protections through routine overrides while preserving the flexibility needed for genuine edge cases.

3.2.6 Non-Circularity Demonstration

These definitions are not circular because they separate components with distinct roles and independent specifications:

Component	Defined By	Independent Of
Impacts	Welfare indicators, calibration (§5), ripple propagation (§8)	Rights thresholds, tail-risk bounds
Admissibility	Rights floors (§6), tail-risk corridor (§7)	RLS scoring weights
Ranking (≻)	RLS/UCI/HOI (§11) applied to admissible set only	Admissibility criteria
Containment	UCI change thresholds (§3.4, §11.6)	Admissibility criteria
Good/Bad labels	Derived from admissibility + containment + ranking	—

No term in this chain depends on a later term for its definition.

3.2.7 Canonical Impact Construction Algorithm

Baseline-Zero Rule (Normative). For all cells (u,d) and all indicators used to construct impacts, the impact is defined as a change from baseline:

I_{u,d}(a) = 0 if and only if the predicted indicator state under option a equals the baseline indicator state, under the same mapping and anchors.

Any example that computes a level score S(x) MUST immediately difference it against baseline S(x_0) to produce a delta impact.

To eliminate ambiguity in how impacts are computed, MathGov specifies the following canonical algorithm. This algorithm is the single authoritative pipeline; alternative methods (such as direct percentile anchoring without instance decomposition) are special cases that must map into this structure.

Algorithm: Canonical Impact Construction

Input: Decision context, option a, welfare indicator data, kernel profile

Output: Propagated, saturated impact matrix Ī^prop_{u,d}(a) for all (u, d)

Step 1: Indicator Selection and Measurement

For each active cell (u, d) where m_{u,d} = 1:

1. Select welfare indicators consistent with dimension definitions (Section 5.1)
2. Measure or estimate indicator values under baseline and under option a
3. Record data sources, measurement protocols, and uncertainty bounds in PCC

Step 2: Magnitude derivation (μ_k) (baseline-delta canonical).

Define a reference-class scoring function S(·) mapping the raw indicator x into a bounded score in [−1, +1]. Default (percentile-anchored, higher-is-better):

[S(x) := clip((x − P50) / (P95 − P5) , −1 , +1).]

If higher values are worse, use S_worse(x) := −S(x) (or equivalently apply a monotone sign correction).

Canonical impact meaning: impacts are changes from baseline, so the magnitude used for an instance is:

[μ_k := clip(S(x_a) − S(x_0) , −1 , +1),]

where x_0 is the baseline indicator value and x_a is the value under option a.

Notes. This preserves the paper's global meaning that 0 = no change, while still allowing percentile anchoring to define scale sensitivity.

For purely change-based indicators already expressed as Δx, analysts MAY compute μ_k directly from Δx using a declared change reference class, but the PCC must still show that μ_k = 0 when x_a = x_0.

Rights-anchor override (mandatory for rights-covered cells). For any rights-covered cell, percentile anchoring using P5/P50/P95 MUST NOT be used unless the PCC explicitly declares the reference class as Invariant and links it to the Rights Anchor Registry (Appendix T / REG-RIGHTS-ANCHORS-*). In the default case for rights-covered cells, μ_k MUST be derived from the applicable invariant rights anchor mapping specified in Appendix T.

Step 3: Instance Aggregation

Audit requirement. The PCC must list all impact instances k contributing to each (u,d) cell, including the declared evidence class, uncertainty/range parameters, and any decomposition/aggregation mode selection, so the same instance list yields identical cell totals under deterministic replay.

For each cell (u, d), aggregate impact instances:

$$\tilde{I}^{dir,pre}_{u,d}(a) = \Sigma_{k \in K_{u,d}} [r_k \cdot \tau(t_k) \cdot \ell_k \cdot c_k \cdot e_k \cdot \mu_k]$$

Where $K_{u,d}$ is the set of impact pathways/instances mapped to cell (u,d). (All multiplicative factors apply within an instance k; instances sum additively.)

where r_k is reach, $\tau(t_k)$ is temporal weight, ℓ_k is likelihood, c_k is confidence, and e_k is equity adjustment.

Authoritative aggregation rule. Unless the PCC explicitly declares an alternative, MathGov uses the canonical weighted-sum formula in Appendix B.2.4. Each instance contribution is computed as the product of its attributes (reach r_k, temporal weight $\tau(t_k)$, likelihood ℓ_k, equity/resilience adjustment e_k, confidence c_k, and magnitude μ_k), and the cell total is the sum across instances. The key design choice is that confidence c_k enters multiplicatively within each instance contribution, so low-confidence claims contribute proportionally less to the aggregate, which reduces both expected impact and downstream propagated effects.

Terminology note. The aggregation is additive across instances (a sum). The term "multiplicative" refers to how the instance attributes multiply within each summed term, not to multiplying terms together across instances.

Step 4: Direct Impact Saturation

$$I^{dir}_{u,d}(a) = sat_\beta(\tilde{I}^{dir,pre}_{u,d}(a)) , \text{ with } sat_\beta(x) := \tanh(\beta \cdot x)$$

Note: $sat_\beta(x)=\tanh(\beta \cdot x)$ is asymptotic at ±1. The framework treats [-1,+1] as representational bounds; implementations MUST NOT clip unless a specific rule explicitly calls for clipping.

Interpretation: β controls the curvature of diminishing returns; larger β saturates faster. Default: $\beta = 2$ (unless the PCC declares a different β and justifies it).

Step 5: Vectorization

Flatten the 7 × 7 direct impact matrix into a 49-element vector $\mathbf{I}^{dir}$ using the canonical flattening map $\phi(u, d) = 7(u - 1) + d$.

Indexing note: the flattening map $\phi(u,d)$ is defined 1-based (outputs in $\{1,\ldots,49\}$). If implementing in a 0-based language, use $\phi0(u,d) := \phi(u,d) - 1$, and apply the same shift consistently to any kernel/vector indices.

Step 6: Ripple Propagation

Apply kernel propagation:

Quick mode:

$$vec(\tilde{I}^{prop,pre}(a)) = vec(I^{dir}(a)) + K \cdot vec(I^{dir}(a))$$

(First-order approximation; uses the declared kernel K and the canonical flattening vec(·) defined in Step 5.)

*Full mode (requires ρ(**K**) < 1):*

$$\mathrm{vec}(\tilde{I}^{prop,pre}(a)) = (I - K)^{-1} \cdot \mathrm{vec}(I^{dir}(a)) = \Sigma_{n=0}^{\infty} K^n \cdot \mathrm{vec}(I^{dir}(a))$$

(Full resummation; valid only when the stability/invertibility conditions for (I−K) are satisfied and recorded in the PCC.)

Step 7: Post-Propagation Saturation

$\bar{I}^{prop}_{u,d}(a) = \mathrm{sat}_{\beta_{prop}}(\tilde{I}^{prop,pre}_{u,d}(a))$, with $\mathrm{sat}_{\beta_{prop}}(x) := \tanh(\beta_{prop} \cdot x)$

Default: $\beta_{prop} = 1$ (unless the PCC declares otherwise).

Step 8: Scenario Conditioning (if applicable)

For scenario-aware analysis (and bounded-impact diagnostics where permitted), repeat Steps 1-7 for each scenario s to obtain $\bar{I}^{prop}_{u,d}(a \mid s)$.

Step 9: Documentation

Record in PCC: indicator definitions, reference classes, anchoring method, instance attributes, kernel profile, propagation mode, and all intermediate values.

3.2.8 Distributional Rights Semantics

A critical protection against rights violations being averaged away within heterogeneous populations:

Principle: For NCRC-covered cells, rights checks must be applied to the worst-off subgroups, not merely to aggregate or mean impacts.

Implementation (worst-off operator). For any rights-covered cell (u, d), let $G_{u,d}$ be the set of protected/vulnerable subgroups identified under §3.2.9. Let $\bar{I}^{prop}_{u,d}(a \mid g)$ be the post-propagation, post-saturation impact for subgroup g. Define the rights-check impact used by NCRC as:

$$[\ \bar{I}^{rights}_{u,d}(a) := \min_{g \in G_{u,d}} \bar{I}^{prop}_{u,d}(a \mid g).\ ]$$

If $G_{u,d}$ is empty due to infeasible subgroup enumeration, invoke the Tier-gated fallback in §3.2.9 (“Unknown Subgroup Trigger”).

3.2.9 Subgroup Analysis Protocol

For all rights-covered cells, subgroup analysis follows this canonical order:

Step 1: Subgroup Identification

For each rights-covered cell (u, d), identify the set of protected subgroups G_{u,d}. At minimum, consider:

- Demographic groups defined by legally protected characteristics

- Economically vulnerable populations (bottom income quintile)

- Geographically exposed communities

- Groups with pre-existing disadvantages in dimension d

Construction rule (to reduce implementer discretion): build G_{u,d} as a deterministic partition of the stakeholder population relevant to cell (u,d). (i) Choose subgroup axes A = [a1,…,am] (at minimum: one legally protected axis plus any context-critical axes declared in the PCC). (ii) For each axis aj, take its subgroup labels from the run's data dictionary (no ad hoc relabeling). (iii) Form candidate groups as all single-axis groups; optionally add intersections of the two highest-risk axes declared in the PCC. (iv) Enforce a minimum subgroup size n_min (default n_min = 30 unless the domain dataset is smaller; if smaller, set n_min := max(5, ⌈0.05·N⌉)). (v) Any candidate group with n < n_min MUST be merged into a deterministic 'Other/Small' bucket for that axis-set. (vi) After merging, require |G_{u,d}| ≥ 2; if not, set G_{u,d} := {All, Other/Unknown} and create a phantom 'Unknown' subgroup bound per §3.2.8. All axis choices, n_min, and any intersections used MUST be recorded in the PCC so an independent challenger can reproduce the same partition.

Tier-3 starter suggestion: γ_subgroup := 1.5. (If used, Tier-3 releases SHOULD declare a value in the PCC; Tier-4 ignores γ_subgroup in favor of explicit subgroup enumeration.)

Define a conservative rights-check impact when only aggregate impacts are available:
Ī^{rights}_{u,d}(a) := max(−1, γ_subgroup · Ī^{prop}_{u,d}(a)) when Ī^{prop}_{u,d}(a) < 0, and
Ī^{rights}_{u,d}(a) := Ī^{prop}_{u,d}(a) otherwise.

If enumeration is infeasible (Tier-3 conservative bound). When only aggregate impacts are available and subgroup enumeration is infeasible, implementers MAY apply γ_subgroup only to negative aggregate impacts for rights checking (not for RLS ranking), as a conservative approximation.

Minimum subgroup categories (Tier-3). For rights-covered cells, consider at minimum:
- Demographic: age cohorts (children under 18, elderly 65+), disability status, and other protected classes as locally relevant;
- Economic: income quintiles (especially bottom quintile), housing insecurity, employment status;
- Geographic: urban/rural divide, regions with known service disparities;
- Domain-specific: stakeholder groups with asymmetric exposure to the decision.

Tier-4 Subgroup Enumeration Policy (Normative)

To prevent implementer forks, Tier-4 Pilot-Executable runs MUST treat subgroup disaggregation as an explicit PCC input, not an implicit choice.

Tier-4 requirements:

1) Rights-covered cells. For every (u,d) cell that is referenced by any rights check (NCRC or Emergency Mode), the PCC MUST declare a finite subgroup set $G_{u,d}$. The PCC MUST report subgroup-conditioned propagated impacts $\bar{I}^{prop}_{u,d}(a|g)$ for each candidate option a and each $g \in G_{u,d}$.

2) Minimum subgroup count. Unless union u is single-entity for this decision (definition below), $|G_{u,d}|$ MUST be ≥ 2 for every rights-covered cell.
Single-entity exemption (normative). Union u is single-entity for this decision iff: (i) exactly one stakeholder instance exists in union u for this decision's scope, (ii) no meaningful intra-union subgrouping exists that could change rights exposure for this (u,d) cell, and (iii) the PCC records a brief justification.
If this exemption is used, set audit_flag = SUBGROUP_SINGLE_ENTITY_EXEMPTION_USED and record the justification in the PCC.

3) Worst-off rule is non-negotiable. Rights admissibility MUST use worst-off subgroup impact: $\min_{g \in G_{u,d}} \bar{I}^{prop}_{u,d}(a|g)$. Subgroup weights do not apply to the minimum.

4) Explicitness over reconstruction. Replay implementations MUST NOT infer subgroup impacts from microdata unless the microdata is hash-bound and explicitly declared in the PCC. Tier-4 replay uses PCC-declared subgroup aggregates only.

Tier-4 note. If subgroup disaggregation is infeasible for a rights-covered cell, the run MUST either (i) declare a conservative subgrouping sufficient to ensure worst-off protection, or (ii) accept a Tier-4 conformance failure for that run (i.e., Tier downgrade).

Step 2: Subgroup-Specific Direct Impact Estimation

For each subgroup $g \in G_{u,d}$, compute direct impacts using the canonical instance pipeline:

$$\tilde{I}^{dir,pre}_{u,d}(a \mid g) = \Sigma_{k \in K(u,d,a,g)} r_k \cdot \tau(t_k) \cdot \ell_k \cdot c_k \cdot e_k \cdot \mu_k, \quad I^{dir}_{u,d}(a \mid g) = \tanh(\beta \cdot \tilde{I}^{dir,pre}_{u,d}(a \mid g)).$$

Step 3: Subgroup-Specific Propagation

Propagate per subgroup (Quick or Full as declared). (Tier 4 Pilot-Executable rev14.x: Quick only; see §13.8.2.)

$$\mathrm{vec}(\tilde{I}^{prop,pre}(a \mid g)) = \mathrm{vec}(I^{dir}(a \mid g)) + K \cdot \mathrm{vec}(I^{dir}(a \mid g)) \quad \text{(Quick)}$$

or

$$\mathrm{vec}(\tilde{I}^{prop,pre}(a \mid g)) = (I - K)^{-1} \cdot \mathrm{vec}(I^{dir}(a \mid g)) \quad \text{(Full; stability-gated).}$$

Then saturate elementwise: $\tilde{I}^{prop}(a \mid g) = \tanh(\beta_{prop} \cdot \tilde{I}^{prop,pre}(a \mid g))$.

Step 4: Worst-Off Identification

[$\bar{I}^{rights}_{u,d}(a) := \min_{g \in G_{u,d}} \bar{I}^{prop}_{u,d}(a \mid g)$.]

3.3 Why "Help" Tracks "Better" and "Harm" Tracks "Worse" (Under the MNA)

Given the MNA, MathGov's use of help and harm as tracking indicators for better and worse is justified as an operational bridge from lived value to computable procedure. It is not asserted as a metaphysically necessary truth, but as the most coherent operationalization under the stated premises.

3.3.1 Phenomenological Grounding (Moral Patient Reality)

For sentient beings, suffering is intrinsically aversive and flourishing intrinsically attractive. This is constitutive of what pain, fear, relief, and fulfillment are like. Under the MNA, this phenomenology provides the core motivational reason to treat reductions in suffering and increases in flourishing as decision-relevant.

In MathGov terms: if $\bar{I}^{prop}_{u,d}(a)$ reliably tracks shifts in welfare-relevant conditions for sentient beings in union *u* along dimension *d*, then positive shifts (Help) supply prima facie reasons, within admissibility constraints, to prefer option *a* over alternatives, while negative shifts (Harm) supply reasons against.

3.3.2 Viability and Self-Defeat Avoidance (Union Persistence Under UBR)

Under UBR, unions are nested and fate-coupled: persistent degradation of containing unions (e.g., biosphere destabilization, societal collapse) eventually collapses the possibility space for flourishing in sub-unions (self, household, community). Systems that systematically erode their own enabling conditions tend toward instability or collapse.

This is not an "ought from is." The structure is:

1. The ought is supplied by the MNA ("preserve conditions enabling flourishing").
2. The is (UBR's interdependence structure) identifies the pathways by which enabling conditions are preserved or destroyed.

Therefore, systematically tracking help/harm across unions and respecting their nesting relationships is instrumentally necessary for any agent committed to the MNA.

Relationship to viability and cybernetic stability traditions. This "self-defeat avoidance" framing aligns with viability theory, which formalizes the conditions under which dynamical systems persist within constraint sets (Aubin, 1991, 2009), and with cybernetic accounts of organizational viability and control (Beer, 1972, 1979). In this reading, UBE can be interpreted as a normative engineering layer placed atop descriptive system viability: once the MNA is adopted, maintaining the enabling conditions for continued flourishing becomes a constraint satisfaction problem over nested unions.

3.3.3 Network Effects and Ripple Coherence (Cooperation Dividends vs. Conflict Spirals)

In interconnected systems, help and harm propagate: cooperation generates trust, knowledge spillovers, and reduced conflict, while harm erodes legitimacy, triggers retaliation, and degrades shared infrastructure. These dynamics are empirically tractable in many domains (Axelrod, 1984; Nowak, 2006) and are represented explicitly in MathGov via the ripple kernel **K** (Section 8).

Conclusion for 3.3: Under the MNA, and given union-structured interdependence, it is coherent to treat helping effects as pro tanto reasons to prefer an option and harming effects as pro tanto reasons to avoid it. This treatment is always subject to non-compensatory admissibility constraints (NCRC/TRC) and structural integrity safeguards (UCI/HOI). The tracking relationship between help and better, and harm and worse, is not primitive. It follows from the MNA combined with UBR's structural claims.

3.4 The Containment Principle (Preventing Pathological Sub-Union "Benefits")

A core failure mode in optimization is local gain purchased by degrading the conditions of the containing system: profit by poisoning the commons, organizational success by eroding social trust, or tumor growth by consuming the host. MathGov addresses this with a containment rule that prevents "growth-by-cannibalizing-context" from being scored as system-level benefit.

3.4.1 Principle Statement

Containment Principle (Normative). Positive impacts on a sub-union do not count as system-level improvements if they materially degrade the coherence or viability of any containing union beyond tolerance.

This is a safeguard against rewarding sub-systems that grow by undermining the larger systems on which they depend. It operationalizes the theoretical distinction between constructive and pathological unions (Section 2.5).

3.4.2 Operationalization (Coherence-Based Check)

Let *u* be a sub-union that shows positive impacts under option *a*. Let Anc(u, D_c) denote the set of containing unions in the nesting hierarchy up to governed depth D_c (default D_c = 2; Section 11.6). Let ΔUCI_j(a) be the predicted coherence shift for containing union *j* under option *a* (computed per Section 11.5).

The containment check for union *u* requires:

where τ_c is a governed tolerance threshold (default τ_c = −0.10; Section 11.6).

Plaintext rendering (audit): For a positive-impact union u, define Containment_u(a) := [min_{u′ ∈ Anc(u, D_c)} ΔUCI_{u′}(a) ≥ τ_c].

Global containment predicate. Define the set of unions with materially positive impacts using the canonical aggregation:

where θ_pos is a governed threshold (default θ_pos = 0.05; Section 11.6) and v_d are the dimension weights from HDW. Then:

Plaintext rendering (audit): Define U_pos(a) := { u : Σ_d v_d · Ī^prop_{u,d}(a) ≥ θ_pos }. Then Containment(a) := ∀u ∈ U_pos(a), Containment_u(a) = true.

Interpretation: If an option that helps a sub-union causes the UCI of any containing union to drop by more than 0.10 (i.e., ΔUCI_j < −0.10), the containment check fails.

Threshold governance: τ_c can be tightened (made closer to zero or positive) for critical containing unions (e.g., Biosphere) but cannot be loosened below the global default without Charter revision. The parameters τ_c, θ_pos, and D_c are PCC governance parameters defined in Section 11.6 and Appendix A.

Note on aggregation consistency: The containment trigger uses weighted aggregation Σ_d v_d · Ī^prop_{u,d} rather than unweighted sum to maintain consistency with RLS aggregation. This ensures that a union is flagged as "positively impacted" under the same weighting scheme used for welfare ranking.

3.4.3 Effect on Evaluation (Governance Modes)

Containment is enforced in two governance-approved modes. The PCC must state which mode is in force:

Mode A (Default for all Tier 4 decisions): Veto / Escalation. If any relevant containing union violates the tolerance (ΔUCI_j < τ_c), option a is flagged as containment-violating and must be either rejected outright or escalated to a higher governance tier for review. This mode prevents containment failures from being silently traded off against other benefits.

Mode B (Exploratory analysis only, Tier 2 or with explicit governance approval): Disqualification of credited gains. If containment fails due to sub-union u, the positive impact contributions from u are disqualified before computing RLS. Formally, replace Ī^prop_{u,d}(a) with min(Ī^prop_{u,d}(a), 0) for all d before computing RLS. This ensures that the sub-union's upside is not credited while any downside is still counted. The option is re-scored accordingly, and the containment failure is prominently recorded in the PCC.

Critical constraint (Normative): Mode B is diagnostic-only and non-binding. Mode B MAY be used only to explore the decision landscape ("what would scores look like if we refused to credit gains from containment-failing sub-unions?"). Mode B outputs MUST NOT be used to determine the final selected option, tie-break outcomes, escalation outcomes, or any admissibility/selection claim.

Selection rule: All binding selection decisions MUST be computed under Mode A (veto/escalation containment gate), as enforced by the canonical selection algorithm (§11.6.2).

Audit enforcement: If a PCC shows Mode B influenced selection, the PCC MUST be labeled INVALID with audit_flag CONTAINMENT_MODE_B_USED_FOR_SELECTION.

This closes the Mode B loophole where disqualification changes which option wins while still claiming Mode B was "exploratory."

Tier restriction: Mode B is prohibited for Tier 4 decisions unless explicitly authorized by governance body with documented justification and mandatory follow-up review within 90 days.

3.4.4 Relationship to NCRC/TRC and the Lexicographic Cascade

Containment is conceptually distinct from NCRC and TRC:

Constraint	Function
NCRC	Protects explicit rights floors (individual and collective)
TRC	Bounds catastrophic tail-risk corridors
Containment	Prevents local optimization from counting as global improvement when it degrades structural viability

In the canonical workflow (Section 11), containment is enforced as an integrity check during selection among admissible options. This occurs after NCRC and TRC have already filtered inadmissible options, but before final RLS-based ranking determines selection.

Cascade placement: Containment is not part of admissibility unless the PCC explicitly elevates it to a veto rule; by default it is enforced as an integrity gate applied to admissible options before final selection. This prevents containment from overriding the lexicographic priority of rights and tail-risk constraints while still providing structural protection.

Relationship to pathological unions: The Containment Principle operationalizes the theoretical distinction between constructive and pathological unions (Section 2.5). An action that "helps" a pathological union (one whose growth degrades containing systems) will fail the containment check and will not be credited as system-level improvement.

3.5 Section 3 Logic Flow Summary

The following diagram illustrates the complete normative logic flow from the MNA through final selection:

MNA (Normative Axiom)

↓

```
Defines: "flourishing matters, suffering should be reduced,
     enabling conditions preserved"
    ↓
Requires operationalization via:
    ↓
┌──────────────────────────────────────────────────────────────────────
──────────────┐
│ 3.2.1 IMPACTS (descriptive)                              │
│  Ī^prop_{u,d}(a) ∈ [-1,+1] — what the option does to welfare        │
│  Canonical Algorithm: indicators → μ_k → instances → saturation →      │
│                  propagation → post-saturation (§3.2.7)          │
└──────────────────────────────────────────────────────────────────────
──────────────┘
    ↓
┌──────────────────────────────────────────────────────────────────────
──────────────┐
│ 3.2.2 HELP/HARM (derived magnitudes)                           │
│  Help = (I)⁺, Harm = (-I)⁺ — sign-consistent decomposition           │
└──────────────────────────────────────────────────────────────────────
──────────────┘
    ↓
┌──────────────────────────────────────────────────────────────────────
──────────────┐
│ 3.2.3 ADMISSIBILITY (deontic filter) — LEXICOGRAPHICALLY PRIOR         │
│  Admissible(a) = NCRC(a) ∧ TRC(a)                         │
│  Forbidden options eliminated BEFORE comparative ranking            │
│  NCRC uses worst-off subgroup impacts (§3.2.8)                 │
│  If A_NCRC = ∅ → Emergency Mode (§6.4)                       │
│  If A_NCRC ≠ ∅ but A_adm = ∅ → TRC Fallback (§7.5)                    │
└──────────────────────────────────────────────────────────────────────
──────────────┘

    ↓

┌──────────────────────────────────────────────────────────────────────
──────────────┐
│ 3.4 CONTAINMENT (integrity gate among admissible options)              │
│  Local Help ≠ System-level Good if it degrades containing unions       │
│  Operationalized via ΔUCI check with τ_c threshold                     │
│  Mode A (veto) vs. Mode B (disqualification, exploratory only)         │
│  Mode B cannot enable selection of containment-violating options       │
└──────────────────────────────────────────────────────────────────────
──────────────┘

    ↓

┌──────────────────────────────────────────────────────────────────────
──────────────┐
│ 3.2.4 RANKING (comparative, among admissible + containment-passing)    │
│  a ≻ b when RLS(a) > RLS(b), with UCI/HOI tie-breaks                   │
└──────────────────────────────────────────────────────────────────────
──────────────┘

    ↓

┌──────────────────────────────────────────────────────────────────────
──────────────┐
│ 3.2.5 GOOD/BAD (derived labels, not primitives)                        │
│  Forbidden-bad, Permitted-better, Permitted-worse, Selected=Good       │
└──────────────────────────────────────────────────────────────────────
──────────────┘

    ↓
```

```
┌──────────────────────────────────────────────────────────────────────
──────────┐
│ 3.3 JUSTIFICATION: Why Help tracks Better, Harm tracks Worse          │
│   (i) Phenomenological, (ii) Viability, (iii) Network effects         │
└──────────────────────────────────────────────────────────────────────
──────────┘
```

4. System Overview: The MathGov Architecture

4.1 The 7×7 Welfare Space (Full Specification)

At the core of MathGov lies a 49-cell welfare matrix formed by the Cartesian product of seven operational unions and seven welfare dimensions.

Unions (rows) U: Self, Household, Community, Organization, Polity, Humanity/CMIU, Biosphere.

Dimensions (columns) D: Material, Health, Social, Knowledge, Agency, Meaning, Environment.

For any candidate action (option) *a*, MathGov represents its welfare consequences as a matrix of normalized impacts:

4.1.1 Normalized Impact Scale

Each cell impact is defined on a bounded, unitless scale:

where:

1. −1 denotes the worst plausible degradation in cell (u, d) under the decision context and calibration protocol,
2. 0 denotes no change relative to the baseline,
3. +1 denotes the best plausible improvement in that cell under the same calibration.

Impacts are context-calibrated (Section 5.4): the mapping from real-world indicators to [−1, +1] depends on the decision scope, reference population/system, time horizon, and anchor datasets.

4.1.2 Intervals and Epistemic Humility

When uncertainty is material, MathGov records impacts as intervals:

with an associated confidence/provenance record in the PCC. Interval endpoints are epistemic bounds (what the decision process considers plausible), not frequentist confidence intervals unless explicitly stated.

4.1.3 Direct vs. Propagated Impacts (Pipeline)

MathGov distinguishes three related impact objects:

Direct impacts (before ripple propagation), produced from impact instances and saturated:

Propagated (pre-saturation) impacts after ripple propagation:

or

where **I**^dir is the flattened direct-impact vector, **K** is the sparse ripple kernel, and $\mathbf{I}_{49}$ is the 49 × 49 identity matrix (Section 8.3).

Propagated, post-saturation impacts used for constraints and scoring:

Canonical rule (tiered). NCRC checks use Ī^rights derived from Ī^prop (worst-off subgroup). TRC checks use L_raw(a,s) from AF-BASE/AF-EXT when Tier ≥ 4 (trc_mode = raw_indicator); bounded-impact TRC using Ī^prop(a|s) is permitted only for Tier ≤ 3 as declared, and is diagnostic-only for Tier ≥ 4. RLS ranking uses Ī^prop.

4.1.4 Scenario-Conditioned Impacts (for Scenario-Aware RLS and Diagnostics)

For scenario-aware evaluation (and bounded-impact diagnostics where permitted), impacts are scenario-conditioned. For each scenario s, compute:

and use the scenario-weighted expectation:

where p_s ≥ 0 and Σ_s p_s = 1 (Sections 7.3-7.4). This expectation is used in RLS when scenario evaluation is enabled (Section 11.1; Appendix B).

4.1.5 Applicability Mask (Operational Relevance by Context)

Not every union-dimension cell is operationally meaningful in every decision. MathGov therefore uses an applicability mask:

1. Default (Tier-4): m_{u,d} = 1 for all cells (unless the PCC declares otherwise; see Appendix AD registry).
2. Any m_{u,d} = 0 must be explicitly justified in the PCC (e.g., "Biosphere-Agency not meaningful for this analysis"), and should be used sparingly to prevent metric gaming by omission.

The applicability mask affects RLS aggregation (Section 11.1) but does not override rights or tail-risk protections: if a cell is relevant to a right or to C_cat, it must not be masked out.

Non-Maskable Cell Enumeration (Always-In-Scope):

The following cell families cannot be excluded via the applicability mask regardless of decision context:

(a) **Rights coverage cells:** All cells (u, d) ∈ C_r for any right r (see Appendix C for the complete mapping).

(b) **Catastrophe cells:** All cells (u, d) ∈ C_cat (see Section 7.2).

(c) **Governance-defined minimum coverage:** At minimum, the following unions must have at least one active cell: Self (U_1), the primary affected union(s), and Biosphere (U_7) for decisions with environmental implications.

The PCC must include a "Non-Maskable Cell Verification" section confirming that all always-in-scope cells have m_{u,d} = 1.

Tier ≥ 4 audit rule: if any rights-coverage cell (u,d) ∈ C_r has m_{u,d} = 0, the PCC is invalid and the run MUST fail with AUDIT_FLAG = RIGHTS_CELL_MASKED_INVALID.

Audit requirement: For any cell (u, d) that is adjacent to a non-maskable cell via the kernel (i.e., K_{ϕ(u,d), ϕ(u',d')} ≠ 0 where (u', d') corresponds to a non-maskable cell), masking requires additional justification explaining why the ripple pathway does not materially affect the non-maskable cell.

4.2 The Lexicographic Cascade

MathGov is a decision methodology designed to avoid three common failures in governance and alignment systems: (i) value scalarization that allows unacceptable tradeoffs, (ii) tail-risk blindness that treats catastrophic downside as “just another term,” and (iii) specification gaming through hidden assumptions or tunable weights. To prevent these failures, MathGov uses a **lexicographic cascade** in which certain constraints are applied as **admissibility filters** prior to any weighted scoring.

The canonical cascade is:

1. **NCRC (Non-Compensatory Rights Constraint)**: remove any option that violates non-negotiable rights floors.
2. **TRC (Tail-Risk Constraint)**: remove any option with unacceptable catastrophic tail-risk.
3. **Containment**: reject or escalate options that create structural fragility or coherence collapse beyond governed limits, even if they pass NCRC and TRC.
4. **RLS (Ripple Logic Score)**: rank remaining admissible options using a weighted welfare aggregation across unions and dimensions.
5. **UCI/HOI tie-breaks**: when RLS differences are not decisively separable, compare options using structural coherence metrics (UCI) and hollowing diagnostics (HOI), then apply governance rules for judgment calls and escalation.

This ordering implements the core principle of non-compensability: **no welfare gain may override a rights violation or unacceptable catastrophic risk**, and no high-scoring option may be selected if it fails containment.

4.2.1 Canonical decision flow (algorithmic statement)

Let Obe the set of candidate options. Define the admissible sets:

1. $A_{NCRC} \subseteq O$: options that satisfy NCRC
2. $A_{adm} \subseteq A_{NCRC}$: options that satisfy both NCRC and TRC

The canonical decision flow is:

START

1. Generate option set O.
2. Apply NCRC to form A_{NCRC}. If $A_{NCRC} = \emptyset$, invoke **NCRC Emergency Mode**.
3. Apply TRC to form $A_{adm} \subseteq A_{NCRC}$. If $A_{adm} = \emptyset$, invoke **TRC Fallback Mode**.
4. Compute $RLS(a)$for all $a \in A_{adm}$(and uncertainty if enabled).
5. Order candidates by $RLS$from best to worst.
6. Apply **Containment (Mode A)** as an integrity gate prior to selection: evaluate containment for the current best candidate; if it fails, reject or escalate per §11.6 and evaluate the next candidate.
7. If the leading candidates are within the discrimination band, apply UCI/HOI tie-break rules per §11.4–§11.6.
8. Generate PCC for the selected or escalated outcome.
   END

Computational note. Containment is conceptually a pre-selection integrity gate. To reduce computation, implementations may evaluate containment only for the leading candidates in descending RLS order, provided no containment-violating option is selected without Mode A escalation and PCC documentation.

Figure 4.2-A: MathGov Decision Pipeline (v5.0i)

The following diagram illustrates the complete decision pipeline from inputs through selection. Each gate is lexicographically prior to subsequent stages, failure at any gate excludes the option from downstream processing (except under explicitly declared emergency or fallback modes).

```
╔══════════════════════════════════════════════════════════════════════
════════════════════════╗
║                MATHGOV DECISION PIPELINE (v5.0i)                    ║
║                  Tier-4 Pilot-Executable Flow                    ║
╠══════════════════════════════════════════════════════════════════════
════════════════════════╣
║                                                      ║
║ ┌────────────────────────────────────────────────────────────────────
──────────┐ ║
║ │                  INPUTS (PCC Header)                   │ ║
║ │                                                  │ ║
║ │ • decision_id, decision_owner, timestamp, spec_version              │ ║
║ │ • scope, unions in scope, dimensions in scope, time horizon          │ ║
║ │ • option_set O = {a1, a2, ..., an}                         │ ║
║ │ • baseline state x0                                  │ ║
║ │ • registry_hashes, {rights_anchors, thresholds, AF-BASE, kernel,       │ ║
║ │              weights, scenario_library}                   │ ║
║ │ • configuration, tier, propagation_mode, trc_mode                 │ ║
║ │                                                  │ ║
║ │ Output, Initialized PCC draft                            │ ║
║ └──────────────────────────────────────────┬─────────────────────────
──────────┘ ║
║                         │                            ║
║                         ▼                            ║
║ ┌────────────────────────────────────────────────────────────────────
──────────┐ ║
```

```
║ │                1. IMPACT ESTIMATION                    │ ║
║ │                                        │ ║
║ │ For each option a ∈ O:                                 │ ║
║ │   • Construct direct impacts I^dir_{u,d}(a) via instance pipeline     │ ║
║ │   • Apply kernel propagation (None or Quick or Full per config; Tier 4 Pilot-Executable rev14.x:
Quick only)     │ ║
║ │   • Apply post-propagation saturation → I^prop_{u,d}(a)            │ ║
║ │   • For rights-covered cells: compute worst-off subgroup impacts      │ ║
║ │     I^rights_{u,d}(a) = min_g I^prop_{u,d}(a|g)                  │ ║
║ │   • For scenario-aware evaluation: compute scenario-conditioned impacts I^prop(a|s) (Tier ≥ 4
TRC uses L_raw(a,s)) │ ║
║ │                                        │ ║
║ │ Output, Impact matrices for all options                     │ ║
║
└──────────────────────────────────────────────┬──────────────────────────────────────────────────────
──────────────┘ ║
║                     │                       ║
║                     ▼                       ║
║
┌─────────────────────────────────────────────────────────────────────────────────────────────────────
──────────────┐ ║
║ │              2. NCRC GATE (Rights Floor)                 │ ║
║ │                 [LEXICOGRAPHIC LEVEL 1]                  │ ║
║ │                                        │ ║
║ │ For each option a ∈ O:                                 │ ║
║ │  For each right r ∈ R:                               │ ║
║ │    For each cell (u,d) ∈ C_r:                           │ ║
║ │      Check: I^rights_{u,d}(a) ≥ θ_r                        │ ║
║ │                                        │ ║
║ │
┌──────────────────────────────────────────────────────────────────
───┐ │ ║
║ │ │ ✗ ANY violation → INADMISSIBLE │ │ ║
║ │ │ │ │ ║
║ │ │ If A_NCRC = ∅ (no option passes): │ │ ║
║ │ │   → Invoke EMERGENCY MODE (see §6.4) │ │ ║
║ │ │   → Lexicographic minimization of violation depths │ │ ║
║ │ │   → Mandatory remediation plan │ │ ║
║ │ │ │ │ ║
║ │ │ ✓ ALL rights satisfied → Option enters A_NCRC │ │ ║
║ │
└──────────────────────────────────────────────────────────────────
───┘ │ ║
║ │ │ ║
║ │ Output, A_NCRC = {a ∈ O : NCRC(a) = true} │ ║
║
└──────────────────────────────┬───────────────────────────────────
────────┘ ║
║          │          ║
║          ▼          ║
║
┌──────────────────────────────────────────────────────────────────
────────┐ ║
║ │        3. TRC GATE (Tail-Risk Constraint)        │ ║
║ │          [LEXICOGRAPHIC LEVEL 2]          │ ║
║ │ │ ║
║ │ For each option a ∈ A_NCRC: │ ║
║ │   • Compute scenario losses L(a,s) for s ∈ S │ ║
║ │   • Compute CVaR_a(L(a)) using discrete algorithm                   │ ║
║ │   • Check: CVaR_a(L(a)) ≤ τ_TRC                          │ ║
║ │                                                    │ ║
║ │ Output, A_adm = {a ∈ A_NCRC : TRC(a) = true}                    │ ║
║
└────────────────────────────────────────────┬──────────────────────────────────────────
─────────────┘ ║
║                     │                      ║
║                     ▼                      ║
║
┌───────────────────────────────────────────────────────────────────────────────────────
─────────────┐ ║
║ │              4. CONTAINMENT GATE (Structural Integrity)              │ ║
║ │                     [INTEGRITY CHECK]                       │ ║
║ │                                                    │ ║
║ │ Uses ΔUCI for containing unions. Mode A is mandatory for selection.    │ ║
║ │                                                    │ ║
║ │ Output, Selectable set                                     │ ║
║
└────────────────────────────────────────────┬──────────────────────────────────────────
─────────────┘ ║
║                     │                      ║
║                     ▼                      ║
║
┌───────────────────────────────────────────────────────────────────────────────────────
─────────────┐ ║
║ │                     5. RLS RANKING                        │ ║
║ │                                                    │ ║
║ │ Compute RLS(a) for selectable options and rank.                     │ ║
║ │                                          │ ║
║ │ Output, Ranked list                          │ ║
║
└───────────────────────────────────────────────────┬───────────────────────────────────────────────
──────────────┘ ║
║                    │                      ║
║                    ▼                      ║
║
┌───────────────────────────────────────────────────────────────────────────────────────────────
──────────────┐ ║
║ │          6. TIE-BREAK (if non-decisive)             │ ║
║ │                                          │ ║
║ │ Apply UCI dominance then HOI risk flag then escalation.          │ ║
║ │                                          │ ║
║ │ Output, Selected option a*                       │ ║
║
└───────────────────────────────────────────────────┬───────────────────────────────────────────────
──────────────┘ ║
║                    │                      ║
║                    ▼                      ║
║
┌───────────────────────────────────────────────────────────────────────────────────────────────
──────────────┐ ║
║ │                7. OUTPUTS                  │ ║
║ │                                          │ ║
║ │ Final PCC contains cascade trace, selection rationale, 5SPR, signatures.│ ║
║ │                                          │ ║
║
└───────────────────────────────────────────────────────────────────────────────────────────────
──────────────┘ ║
```

```
║                                        ║
╠══════════════════════════════════════════════════════════════════════════════════════════════
════════════════════════════════╣
║ LEGEND                                      ║
║ O, option set; A_NCRC, rights-admissible; A_adm, fully admissible.       ║
║ RLS, Ripple Logic Score; UCI, Union Coherence Index; HOI, Hollowing-Out     ║
║ Index; CVaR, Conditional Value-at-Risk; PCC, Provenance and Compliance      ║
║ Certificate; 5SPR, Five-Sentence Public Rationale.                ║
╚══════════════════════════════════════════════════════════════════════════════════════════════
════════════════════════════════╝
```

Figure notes. Lexicographic priority is enforced by the cascade structure. Emergency and fallback modes are invoked only when normal processing yields an empty admissible set. Mode A containment is mandatory for selection, Mode B is diagnostic-only.

4.2.2 Outputs and audit artifacts

Every Tier 4 application produces a PCC (Provenance and Compliance Certificate) that includes:

1. the option set $O$and how it was generated,
2. all NCRC and TRC parameters used (including C_r, C_{cat}, and catastrophe weights),
3. the computed admissible sets $A_{NCRC}$and A_{adm},
4. RLS computations and weights (union weights and dimension weights),
5. containment results and any escalation,
6. tie-break results and judgment-call triggers (if applicable),
7. and declared overrides or deviations from defaults.

This ensures the cascade is independently reproducible and audit-ready.

This section is included to prevent misreads in packaged audits that confuse “parameterization required” with “procedure missing.”

• Tier 4 (high-assurance institutional): requires (i) non-placeholder invariant rights anchors for all active rights (Appendix T registry), and (ii) either a validated kernel library or a domain-calibrated kernel with stated uncertainty bounds and governance sign-off. If either is absent, the decision cannot be claimed Tier 4, and must downgrade or escalate per NCAR.

• Tier 2–3 (pilots and standard audits): admissibility and ranking are executable with conservative defaults, including the explicit option K = 0 (no propagation) when kernel evidence is insufficient, and rights thresholds/anchors as specified in Appendix T with PCC disclosure.

Tier-specific expectations (default, conservative):

• Deployment-calibrated: the chosen anchors, indicators, and propagation kernel are empirically grounded for the target domain (a Tier-dependent requirement).

• Methodology-complete: the decision procedure is executable given declared inputs and governed parameters (true in this document).

For audit clarity, MathGov distinguishes:

MathGov is a complete methodology specification: the lexicographic cascade is fully defined, and every admissibility or ranking step has a formally specified predicate or computation. However, some numerical objects are intentionally governed, domain-instantiated parameters. This is not a gap in the decision logic; it is the boundary between a universal method and a context-specific deployment.

4.2.3 Completeness and instantiation boundary (clarification)

4.3 Inputs, representations, and comparability requirements

MathGov compares options by representing their consequences in a common decision structure. For each option a, the methodology requires:

1. **Declared decision context**: the scope, stakeholders, time horizon, and decision tier.

2. **Union and dimension scope**: the unions and welfare dimensions included, using the canonical ordering unless explicitly overridden.

3. **Impact representation**: direct impacts are represented in the 7×7 union–dimension matrix with bounded impact values in $[-1, +1]$and with explicit reference classes, anchors, and sign conventions.

4. Uncertainty representation (optional but recommended at Tier 4): scenario sets, probabilities, and the uncertainty model used for TRC and for any uncertainty-adjusted discrimination threshold.

5. **Governance parameters**: union weights, dimension weights, catastrophe cell set C_{cat}, catastrophe weights ω, rights thresholds θ_r, tail-risk thresholds, containment limits, and any overrides.

4.3.1 Comparability rule across options

All compared options must be evaluated under:

1. the same union and dimension set,

2. the same impact scaling and anchoring conventions,
3. the same rights coverage sets and thresholds,
4. the same catastrophe cell set and catastrophe weight rules,
5. and the same containment policy.

If two options cannot be made comparable, the PCC must either:

1. reject the comparison as invalid, or
2. decompose into comparable sub-decisions and apply the cascade separately.

4.3.2 Tier-based requirements for uncertainty and subgroup analysis

Tier determines minimum methodological requirements:

1. Subgroup analysis is mandatory for Tier 4 decisions and recommended for Tier 3. Tier 1 decisions may use aggregate impacts with explicit PCC acknowledgment of limitation.
2. Scenario modeling for TRC is strongly recommended for Tier 4 and mandatory for Tier 4 when catastrophic risk is nontrivial.

4.4 Implementation tiers and minimum compliance requirements

MathGov is operational at multiple levels of rigor. Tiers define minimum compliance requirements, not optional features. From v4.8.6 onward, tiers are numbered Tier 1 to Tier 4. The prior intermediate tier label is removed; its conservative ripple requirements are incorporated into Tier 3.

Definition (Tier vs. Propagation Mode). MathGov uses two independent “dials”:

1) Implementation Tier (Tier 1–4) sets the minimum compliance and assurance level: what checks are mandatory, what audit artifacts are required, and what strength of claim is permitted.

2) Propagation Mode (None / Quick / Full) sets how ripple coupling through the kernel K is computed, if propagation is used at all:

- None (Direct-only): $\bar{I}^{prop}(a) := I^{dir}(a)$. No kernel propagation.
- Quick mode (first-order): $\bar{I}^{prop}(a) := I^{dir}(a) + K I^{dir}(a)$. This is a conservative, computationally light approximation.
- Full mode (resummed): $\bar{I}^{prop}(a) := (I - K)^{-1} I^{dir}(a)$, used only when stability/invertibility conditions are satisfied and declared.

Tier–mode interoperability (canonical defaults). Tiers do not imply a propagation mode, but they constrain which modes are allowed:

- Tier 1: No propagation requirement (typically None).
- Tier 2: Default None (Direct-only). Quick mode MAY be used if a declared starter kernel is available; Full mode is not permitted.
- Tier 3: Quick mode MAY be used when propagation is claimed; Full mode is not permitted unless explicitly escalated to Tier 4 governance.
- Tier 4 (Pilot-Executable, rev14.x): FULL propagation is prohibited. Allowed propagation_mode values are NONE or QUICK only. Tooling MUST hard-fail any Tier-4 Pilot-Executable run that sets propagation_mode=FULL, and MUST record audit_flag FULL_PROPAGATION_PROHIBITED_TIER4_REV14. Full propagation MAY be described only as a future Tier-4 Certified profile under a new revision with a hash-bound deterministic solver profile and an updated NDP.

To prevent ambiguity, the terms "Quick" and "Full" are reserved for propagation mode only and are not used as tier labels.

Figure 4.4-A. Two independent dials in MathGov: Implementation Tier (governance and minimum compliance) and Propagation Mode (kernel ripple depth). Quick/Full refer only to propagation computation, not tier labels.

Document Scope Boundaries (Normative)

To support independent implementation and audit, MathGov clearly delineates where different types of content reside:

Foundation Paper Scope. This document specifies: the lexicographic cascade and its mathematical definitions; admissibility predicates (NCRC, TRC) and their evaluation rules; containment, scoring, and tie-break logic; governance requirements including HDW, NCAR, and tier policies; PCC structure and audit obligations; and validation criteria. The Foundation Paper defines what must be computed and how compliance is evaluated. It does not embed scenario libraries, kernel edge values, registry data files, or executable code.

Appendices Volume Scope. The companion Appendices document provides: complete symbol and notation references; canonical equations in consolidated form; rights coverage mappings and threshold calibration protocols; UCI indicator families and measurement operationalization; TRC parameter defaults and scenario governance templates; implementation roadmaps and quick-start guides; and glossary, version history, and cross-reference materials. The Appendices provide definitions and reference structures that instantiate the methodology.

ProofPack Scope. The companion ProofPack artifact package provides the integrity scaffolding needed to make Tier-4 Pilot-Executable claims replayable by third parties (via hash-bound schemas, canonicalization rules, manifests, and registries). This manifest-only release does not ship executable tooling for replay or conformance testing.

Tier-4 note (rev14.1): ProofPack registries are shipped as hash-bound JSON files in the ProofPack bundle; they are not “spec-only for later extraction.”

What "manifest-only" means. A manifest-only release ships a hash-bound ProofPack (data-only) that contains JSON registries, JSON schemas, and MANIFEST index files (plus optional test vectors), but ships no executables. Registries are not extracted from prose. Implementers verify integrity by canonicalizing each JSON artifact under the declared canonicalization profile and checking its sha256 against the ProofPack manifests and the PCC. Any bundle with placeholders remaining in MANIFEST files is NONCONFORMANT as a release artifact.

4.4.1 Tier 1 (Heuristic)

Use when decisions are low-stakes and reversible. Tier 1 supports rapid, qualitative application of Union-Based Ethics (UBE) and Ripple Logic without claiming full calculability.

Minimum requirements: (i) state the decision question and options; (ii) identify affected unions and likely rights risks; (iii) record a brief rationale and any obvious uncertainty; (iv) if any rights risk is plausible, escalate to Tier 2+.

4.4.2 Tier 2 (Core, Calculable)

Use for routine decisions requiring a transparent, calculator-checkable run. Tier 2 implements the full cascade with conservative defaults and minimal modeling burden.

(iii) NCRC is required; TRC is recommended and becomes required when the decision plausibly implicates catastrophe cells C_cat or mandatory tail scenarios (Tier escalation trigger; see authoritative Tier Requirements Matrix §4.4.5).

Normative Hierarchy (Single Source of Truth) (Normative).

When interpreting this specification, the following precedence order applies (highest controls lowest):

1) PCC-bound, hash-referenced registries and embedded snapshots (including ProofPack (manifest-only bundle) for Tier-4 when invoked);

2) Tier Requirements Matrix (§4.4.5);

3) Default Policy by Tier (§4.4.6);

4) Section-level normative rules (MUST/SHALL);

5) Default tables labeled as “Normative defaults unless overridden by PCC registries”;

6) Examples, MREs, and illustrative guidance (non-normative unless explicitly declared).

If any statement conflicts across levels, the higher-precedence level controls, and the PCC SHOULD record audit_flag DOC_PRECEDENCE_CONFLICT when a lower-level default is intentionally overridden.

4.4.3 Tier 3 (Standard, Auditable)

Use for high-stakes or contested decisions where auditability and conservative ripple treatment are required. Tier 3 corresponds to the prior combination of standard and conservative-ripple practices.

Minimum requirements: (i) produce a PCC (Provenance and Compliance Certificate) including configuration, weights, scenario set, and any overrides; (ii) perform NCRC subgroup analysis, or if infeasible, document the limitation and apply the Tier 3 conservative bound defined in Section 6.5; (iii) satisfy TRC under the mandatory-tail scenario policy and red-team requirements where applicable; (iv) if propagation is used, Quick mode is permitted with a declared starter kernel and KQS score, and propagation uncertainty must be included; (v) containment is enforced as a selection gate on admissible options unless explicitly overridden via escalation, with the override logged in the PCC.

Kernel evidence at Tier 3. Class E (elicited) kernel entries are permitted at Tier 3 with mandatory PCC disclosure: “Kernel based on starter KOPS (Class E); not domain-validated.” Sensitivity analysis showing selection stability under ±0.05 perturbation of any relied-upon non-zero entries is REQUIRED when any kernel entry materially affects the outcome.

4.4.4 Tier 4 (High Assurance, Institutional)

Use for institutional deployment, safety-critical contexts, or repeated policy decisions where model governance and evidence-backed propagation are required.

Minimum requirements: (i) all Tier 3 requirements; (ii) all active rights used for NCRC must have non-placeholder Invariant Rights Anchors defined in Appendix T (or an appended registry); if any required anchor is placeholder, downgrade the decision to Tier 3 or require explicit escalation; (iii) scenario governance includes independent elicitation and red-team review for mandatory tails; (iv) Full propagation may be used only when evidence supports kernel structure and stability (including the declared stability condition) and KQS meets the Tier 4 threshold; otherwise enforce conservative propagation (K=0) or Quick mode with declared limitations; (v) governance changes to thresholds, weights, kernel entries, scenario libraries, and anchor registries follow the Charter-controlled versioning protocol and are recorded in the PCC. Not permitted in Tier-4 Pilot-Executable rev14.x; relevant only to a future Tier-4 Certified profile.

Tier mapping note. Legacy references to the prior intermediate tier label in prior drafts are superseded; requirements formerly labeled the prior intermediate tier label are included in Tier 3 in this version.

4.4.5 Tier Requirements Matrix (Authoritative)

Normative authority. This matrix is the single authoritative statement of tier compliance. If any other sentence in this document conflicts with this matrix, this matrix governs. All other tier statements MUST refer to this matrix rather than restating requirements.

Capability /	Tier 1	Tier 2	Tier 3	Tier 4 (Pilot-Executable)	Tier 4 (Certified)

Requirement					
Determinism & reproducibility	Best effort	Required (PCC)	Required (auditable)	Required (end-to-end reproducible by third party)	Required + independent verification
PCC artifact	Optional	Required	Required	Required (full, hash-bound)	Required (public + full; certification bundle)
Registry sufficiency (no invention)	Recommended	Required for claimed values	Required	Required: all numbers needed exist in registries (AIL)	Required + signed governance + independent review
NCRC rights admissibility	Heuristic	Required (declared rights)	Required (subgroups for rights cells)	Required (worst-off subgroups; full snapshot)	Required + anchor evidence upgrades
TRC tail-risk admissibility	Optional	Recommended	Required (scenario set)	Required: trc_mode = raw_indicator using AF-BASE; minimum scenario library	Required: raw_indicator + audited scenario library + packaged validation
Scenario library |S| (minimum)	N/A	≥ 5 (recommended)	≥ 20 (context-governed)	≥ 50 (org AI deployment default; declare exceptions)	≥ 50 + packaged stress tests
Kernel propagation	None/Quick	Quick allowed	Quick only (Full prohibited; if Full is needed, escalate to Tier 4).	Quick only (rev14.x). Full prohibited for Tier 4 Pilot-Executable. If FULL requested, hard-fail Tier 4 claim (FULL_PROPAGATION_PROHIBITED_TIER4_REV14).	Full allowed only when a deterministic solver profile is hash-bound and NDP_FIXEDPOINT_V1 (or later) is declared; otherwise Quick. Relied-upon edges ≥ Class B.

Kernel convention & stability	Recommended	Required if used	Required if used	Required: canonical convention; stability guardrails; perturbation sensitivity	Required + independent kernel review
UCI tie-breakers	Optional	Optional	Recommended	Required when RLS ties or uncertainty band overlaps	Required + validated indicator set
KQS / evidence classes	N/A	Required (declare)	Required (audit)	Required; Class E allowed for pilots with flags; not for certification	Required; no Class E on relied-upon edges
Sensitivity analysis bundle	Optional	Recommended	Required (core)	Required: weights, thresholds, kernel ±ε, scenario perturbations	Required + independent replication

Tier-4 pilot-executable meaning (normative). A Tier-4 (Pilot-Executable) claim means that two independent implementers, provided the same PCC and the same referenced registries and configurations, can reproduce the same admissibility outcomes and the same final selection. The pipeline MUST therefore be spec-complete in the sense of deterministic replay.

It implies: (i) every admissibility gate (NCRC, TRC, containment) is fully specified with explicit evaluation procedures; (ii) all required registries are referenced by SHA-256 and are available to the implementer; (iii) ambiguity forks are closed by a single authoritative default policy; (iv) the PCC captures all degrees of freedom (tier, propagation mode, scenario set selection, registry hashes).

It does NOT require: (i) that this release ships executable code or replay tooling; (ii) that every parameter be empirically calibrated; or (iii) that scenario libraries be complete for all domains. Tier-4 requires determinism and auditability given declared artifacts and configurations. Calibration and scenario coverage are governed by tier policy and domain governance processes.

4.4.6 Default Policy by Tier (Single Source of Truth) (Normative)

Purpose. This section resolves all default/parameter ambiguities by tier. If any other section, appendix, or starter artifact states a conflicting default, this section controls.

Tier 1 (Exploratory). No pilot-executable claim. Minimal documentation; propagation and TRC are optional; weights may be informal.

Tier 2 (Core, Pilot Starter). Defaults allowed: (i) union weights and dimension weights MAY be uniform if not specified; (ii) TRC MAY use bounded_impact or bounded-loss diagnostics; (iii) scenario set size MAY be below Tier 3 minimums. Any use of MRC-v1 parameters is Tier-2 only and MUST be labeled Tier 2 in the PCC.

Tier 3 (Standard, Auditable). Requirements are those in the Tier Requirements Matrix (§4.4.5). Default weighting policy: HDW SHOULD be used when constitutional floors and ballots are available; if not, the PCC MUST declare an explicit interim weighting method and the Tier-3 claim is limited to the declared method.

Tier 4 (High Assurance). A Tier-4 Pilot-Executable run MUST (i) satisfy NCRC, TRC, containment, and ranking as specified; (ii) reference all required registries and specifications by SHA-256 in the PCC; and (iii) be deterministically replayable by an independent implementer using conformant tooling built to the ProofPack specifications. This revision is manifest-only and does not ship executable replay tooling.

Supplement dependency (Normative). Tier-4 Pilot-Executable claims depend on the Foundation Paper plus the Appendices plus ProofPack v1.0 at the referenced revision. The PCC MUST reference those artifacts by hash and record all configuration degrees of freedom. If any required artifact is missing, un-hashed, or inconsistent with the PCC, the Tier-4 claim fails.

Operational note. Tier labels and propagation_mode are independent configuration fields. For Tier-4 Pilot-Executable (rev14.x), propagation_mode MUST be NONE or QUICK. If propagation_mode=FULL is requested, the implementation MUST hard-fail the Tier-4 claim and record audit_flag FULL_PROPAGATION_PROHIBITED_TIER4_REV14. For Tier ≤ 3, if FULL is requested but unsupported in the chosen implementation, it MAY fall back to QUICK or NONE, but MUST record the fallback as an audit limitation in the PCC.

Tier-4 Kit Availability Note (Normative). If a Tier-4 run references "ProofPack (manifest-only bundle): MathGov_Tier4_ProofPack_v1.0", the supplement MUST be publicly retrievable under the stated hashes (AE.2). If ProofPack is not available, the run MUST NOT claim Tier-4 Pilot-Executable; it must downgrade to the highest tier satisfied by the available registries and scenario library.

4.5 Justification and Empirical Anchoring of the Seven Dimensions

The seven welfare dimensions used by MathGov are not arbitrary. They represent a convergence zone across multiple independent research programs in human and social flourishing, including the Capability Approach (Nussbaum, 2011; Sen, 1999), Self-Determination Theory (Ryan & Deci, 2000), and Fundamental Human Needs theory (Max-Neef, 1991). They also align with global frameworks such as the OECD Better Life Index (Durand & Boarini, 2016) and the United Nations Development Programme (UNDP) Human Development Reports.

MathGov Dimension	Primary Function	Correlated Literature Domains
Material	Physical resources, income, infrastructure, material security	Nussbaum "control over one's environment"; OECD income & work; Max-Neef subsistence & protection

MathGov Dimension	Primary Function	Correlated Literature Domains
Health	Physiological and psychological functioning	Nussbaum bodily health; WHO quality-of-life measures
Social	Relationships, belonging, social capital	SDT relatedness; Putnam social capital; social cohesion indices
Knowledge	Learning, information access, cognitive development	SDT competence; Capability Approach "senses, imagination, thought"; UNESCO educational rights
Agency	Autonomy, decision power, capacity to act	SDT autonomy; Sen's "agency freedom"; political participation literature
Meaning	Purpose, identity, coherence, existential orientation	Frankl; existential psychology; positive psychology on meaning in life
Environment	Ecological and built context integrity	Planetary boundaries (Rockström et al.); sustainable development and environmental justice

Each dimension is treated as non-fungible at the rights level: gains in one cannot fully compensate for severe harms in another. This multi-criterion structure is a foundational divergence from scalar utilitarianism and is central to the MathGov architecture.

4.6 UBL and UBG Operational Definitions

UBL (Union-Based Living): The practical habituation of MathGov principles into daily decision-making and life patterns.

Operational components:

1. **Daily Union Scan:** Brief morning reflection on anticipated decisions and affected unions
2. **Rights Reflex:** Automatic mental check for obvious rights concerns before action
3. **Ripple Awareness:** Habitual consideration of second-order effects
4. **NCAR Journaling:** Weekly reflection on decisions made and outcomes observed

Measurable indicators:

1. Self-reported use of Tier 1 heuristic (frequency)
2. Rights-near-miss self-identification rate
3. Coherence trajectory in personal domains (self-assessed UCI)

UBG (Union-Based Governance): The institutional infrastructure for implementing MathGov at organizational and policy scales.

Operational components:

1. **Role definitions:** Analyst, Decision Owner, Auditor, Ombudsperson
2. **Decision rights matrix:** Who can decide at each tier
3. **Escalation paths:** When and how to escalate
4. **Charter amendment mechanics:** Supermajority requirements, review periods
5. **Audit cadence:** Minimum review frequency by tier

Minimal viable governance template (organizations):

1. Designate MathGov Analyst role
2. Establish PCC review process
3. Define escalation authority for containment violations
4. Set NCAR reflection schedule
5. Create parameter governance committee

5. Welfare Space: Unions, Dimensions, and Impact Calibration

5.1 The Seven Welfare Dimensions

MathGov represents welfare in a seven-dimensional space D = {1, 2, 3, 4, 5, 6, 7} intended to be (i) broad enough to capture major components of flourishing and harm, (ii) structurally compatible across union scales, and (iii) implementable with measurable indicators. Let:

1. D_1 = Material
2. D_2 = Health
3. D_3 = Social
4. D_4 = Knowledge

5. D_5 = Agency

6. D_6 = Meaning

7. D_7 = Environment

Each union-dimension cell (u, d) receives an impact estimate Ī^prop_{u,d}(a) for option a (produced via the pipeline in Sections 5.2-5.3 and 8).

5.1.1 Dimension Definitions (Canonical Semantics)

Material (D_1). Access to resources and infrastructure needed for survival and functional participation: income/consumption capacity, housing security, food and energy access, critical physical assets, and essential services reliability.

Health (D_2). Physical and mental functioning: morbidity, mortality risk, disability burden, psychological distress/well-being, and the integrity of health-support systems where relevant.

Social (D_3). Relational integrity: belonging, trust, social support, inclusion/exclusion, social capital, and the stability of cooperative networks.

Knowledge (D_4). Epistemic capacity and information conditions: education, skill development, access to accurate information, ability to learn and update beliefs, and resistance to systematic deception.

Agency (D_5). Capacity for self-determination and effective action: autonomy, voice, political participation where relevant, freedom from coercion, and real capability to influence one's conditions.

Meaning (D_6). Coherence, purpose, identity alignment, existential orientation, and the ability to pursue valued life-projects. (Measurement is treated cautiously due to cross-cultural variance; see 5.1.4.)

Environment (D_7). Integrity and quality of the ecological and built contexts that sustain life and functioning: air/water quality, climate stability, biodiversity/ecosystem services, and built environment safety where applicable.

These are dimension semantics, not fixed indicators. Different contexts can use different indicators so long as they validly measure the same latent construct (Section 5.1.4).

5.1.2 Why Seven Dimensions (Design Rationale)

MathGov uses seven dimensions because:

1. They recur across multiple independent research programs (capabilities, needs, well-being, sustainability).

2. They separate "life-support" (Health, Environment, Material) from "cooperation and coordination" (Social, Knowledge, Agency) and "existential coherence" (Meaning) in a way that supports both rights constraints and welfare ranking.
3. They are sufficiently few for tractability in a 7 × 7 matrix but sufficiently rich to avoid collapsing ethics into a single metric.

This is a computational design choice, not a metaphysical claim that welfare has exactly seven axes.

5.1.3 Cross-Union Interpretation (How a Dimension Applies at Different Scales)

Dimensions are defined so they can be applied across unions $u \in U$, but operational indicators differ by union.

Example: Environment

1. For Self/Household/Community/Organization/Polity, Environment captures local ecological and built conditions relevant to that union's functioning (pollution exposure, housing safety, disaster vulnerability, local ecosystem services).
2. For Humanity/CMIU, Environment captures global habitability conditions for civilization (climate stability impacts on food systems, displacement, transboundary pollution).
3. For Biosphere, Environment captures Earth-system integrity (planetary boundaries, biodiversity, carbon cycle stability).

Example: Agency

1. For Self: autonomy and real capability to act.
2. For Polity: institutional capacity for legitimate governance and civic participation.
3. For Biosphere: Agency is not assumed to be meaningful as a welfare dimension unless explicitly justified (most analyses should set the applicability mask m_{7,5} = 0 for Biosphere-Agency). See Section 4.1.5 on applicability masks and the requirement to justify exclusions/inclusions.

This is why MathGov includes an explicit applicability mask m_{u,d}: the dimension set is canonical, but applicability can be context-governed without pretending every cell is always meaningful.

5.1.4 Measurement, Indicators, and Cross-Cultural Validity

Each dimension is implemented via indicators, which can vary by culture and context. However, any indicator set must satisfy:

1. **Construct validity:** indicators plausibly measure the stated dimension, supported by literature or empirical validation.
2. **Reliability:** measurement noise is characterized (test-retest where applicable).

3. **Cross-cultural comparability (when required):** for global or cross-population comparisons, perform measurement invariance testing (configural/metric and, when feasible, scalar).

Meaning and some aspects of Agency often have higher cross-cultural variability; therefore MathGov recommends:

1. using multiple indicators,
2. reporting uncertainty intervals,
3. and limiting hard-threshold uses of these dimensions unless validity is strong.

5.1.5 Orthogonality Is a Modeling Goal, Not a Rigid Empirical Claim

MathGov treats dimensions as non-fungible at the rights and tail-risk levels: gains in one dimension do not automatically compensate for severe harms in another. At the measurement level, indicators may correlate.

Therefore:

1. We do not require low correlation among observed indicators.
2. We treat persistent high correlations as a prompt to improve measurement, not as immediate evidence the dimension should be removed.

Redundancy review rule (governance trigger): If repeated measurement programs find strong, persistent correlations suggesting redundancy (default trigger $|r| > 0.85$ across multiple populations and measurement regimes), the system triggers a redundancy review. The review may result in: indicator refinement (better separation), re-factorization within a dimension, or (rarely) dimension restructuring via charter revision. Redundancy review does not automatically imply dimensional merger, because correlated indicators can still represent ethically distinct constructs (e.g., Material and Health).

5.1.6 Rights-Level Non-Compensation Across Dimensions (Bridge to NCRC/TRC)

Even though RLS later aggregates across dimensions for admissible options, MathGov's top layers enforce non-compensation:

1. NCRC assigns rights floors to specific dimension-linked protections (e.g., LIFE/HEALTH, NEED/MATERIAL, INFO/KNOWLEDGE).
2. TRC focuses on catastrophe-relevant cells (notably Health and Environment at Humanity/Biosphere scales).

This preserves ethical structure: measurement correlation does not collapse moral protections.

5.1.7 Indicator Anchoring into the [−1, +1] Impact Scale (Forward Pointer)

Each cell impact is eventually expressed on a normalized scale $[-1, +1]$ relative to a baseline and context-specific anchors (Section 5.4). Practically:

1. -1 corresponds to "worst plausible degradation" for that cell in the decision context,

2. 0 corresponds to "no change from baseline,"

3. $+1$ corresponds to "best plausible improvement."

When indicator mappings are uncertain, MathGov uses interval-valued impacts $[\bar{I}^{lo}, \bar{I}^{hi}]$ with confidence scores and records the mapping and uncertainty in the PCC.

5.2 Impact Instances and Direct-Impact Aggregation

Purpose. This section defines how MathGov converts real-world predicted consequences into a **direct impact score** for each active union–dimension cell (u, d), before ripple propagation (Section 8). MathGov represents each option's consequences as a finite set of **impact instances** k, each instance carrying magnitude, reach, time horizon, likelihood, confidence, and (optionally) an equity/resilience adjustment.

5.2.1 Impact instances

For a given option a and a given active cell (u, d), let:

1. $\mathcal{K}(u,d,a)$ be the set of impact instances asserted for that cell under option a.

2. Each instance $k \in \mathcal{K}(u,d,a)$ has attributes:

1. **Magnitude** $\mu_k \in [-1, +1]$
   Signed direction and severity of the welfare change in that cell for that instance, where positive is beneficial and negative is harmful.

2. **Reach** $r_k \in [0, 1]$
   Proportion of the relevant stakeholder population in union u meaningfully affected by that instance (for that cell and dimension).

3. **Time horizon** $t_k \in (0, \infty)$ years
   The approximate duration over which the instance's effect persists at material relevance for the cell.

4. **Conditional likelihood** $\ell_k \in [0, 1]$
   Probability that the instance occurs, conditional on the scenario model in use. If no scenario partition is used, ℓ_k is conditional on the baseline forecast.

5. **Confidence** $c_k \in [0.1, 1]$
   Analyst confidence in the instance specification and its parameterization (data quality, model support, measurement reliability). The lower bound prevents zeroing out impacts while still penalizing weak claims.

6. Default (Tier-4): e_k = 1.0 (no equity adjustment) unless the PCC declares and justifies a different equity factor; see Appendix AD registry.

All instance attributes and their sources must be recorded in the PCC, including any scenario conditioning used for ℓ_k.

Sentience factor (when applicable). When a cell's impacts are over entities with heterogeneous sentience (SGP §9), define a governed sentience multiplier s_k ∈ (0, 1] for instance k. Default s_k = 1 unless the PCC declares a non-human sentience weighting under the Sentience Gradient Protocol.

Then include s_k multiplicatively in the instance contribution:

[Ĩ^{dir,pre}_{u,d}(a) = Σ_{k ∈ K_{u,d}} r_k · τ(t_k) · ℓ_k · c_k · e_k · s_k · μ_k.]

If the cell concerns only full-rights plateau persons, s_k MUST be 1.

Default s_k derivation (when applicable): s_k is only used when the impacted stakeholder set includes non-human animals or non-person AI systems below the full-rights plateau. In that case, set s_k := clamp(SG(entity)/SG_ref, 0, 1), where SG(entity) ∈ [0,1] is the Sentience Gradient score (see §9 and Appendix F) and SG_ref := 1.00 for the reference full-rights plateau person. If SG(entity) has not been evaluated, set s_k := 1.00 and record SGP_UNEVALUATED = true in the PCC together with a plan for post-run evaluation when the decision context warrants it. Under no circumstances may s_k be used to reduce or trade away rights-floor checks for any full-rights plateau person; it is an instance-weighting modifier only for welfare aggregation.

5.2.2 Missing-data rule (Ignorance Penalty)

MathGov prohibits score inflation by omission. For any active cell (u, d) (m_{u,d}=1) where required measurement is missing or no empirical instances are asserted, the PCC must record an "unknown impact" and apply an ignorance-penalty phantom instance k_phi.

Canonical phantom instance parameters (Tier 4 default):

Tier 4 invalidity rule (Normative). If m_{u,d}=1 for any cell and no measured or estimated impact instances exist, the phantom instance MUST be present. Otherwise the PCC is INVALID with audit_flag ACTIVE_CELL_EMPTY_INSTANCE_SET_INVALID.

μ_phi = −0.10

r_phi = 1, t_phi = T_ref (25 years), ℓ_phi = 1, c_phi = 1.00, e_phi = 1

Tiered guidance for μ_phi may be used for stakes calibration (e.g., Tier 3 or high-stakes contexts), but any deviation from the canonical parameters must be declared in the PCC and sensitivity-tested with the penalty disabled.

5.2.3 Equity/resilience adjustment governance (e_k)

If e_k ≠ 1, the PCC must include:

1. The **equity/resilience criterion** invoked (from a maintained Equity Criteria Registry),
2. A justification linking the criterion to the instance and affected stakeholder set,
3. A **counterfactual audit** reporting the cell's result with e_k reset to 1, and
4. A sensitivity test showing whether the final selection changes when all e_k are set to 1.

Values $e_k > 1$ are permitted only for explicit equity improvements or resilience reinforcement and must be capped by registry limits (default recommended cap $e_k \leq 1.5$ unless an organization explicitly adopts a stricter or looser policy). This prevents silent score manipulation.

5.2.4 Temporal weighting (logarithmic horizon scaling)

Impacts are weighted by a temporal function $\tau(t)$ that maps the time horizon of an instance into a dimensionless multiplier relative to a reference horizon T_{ref}.

Let $T_{ref} = 25$ years by default (approximately one human generation). Define:

$$\tau(t) = \ln(1 + t) / \ln(1 + T_{ref})$$

This yields (illustrative): $\tau(1) \approx 0.21$, $\tau(5) \approx 0.56$, $\tau(10) \approx 0.75$, $\tau(25) = 1.00$, and $\tau(50) \approx 1.22$.

Rationale for logarithmic temporal weighting. The logarithmic form is used instead of exponential discounting for three reasons. First, exponential discounting at common rates (3–7% annually) drives long-horizon impacts toward near-zero present value and systematically marginalizes intergenerational effects and existential tail risks that the TRC is designed to detect. Second, the log form respects temporal non-separability: the ethical significance of an impact should depend on magnitude, reach, and duration, not merely calendar distance. Third, the log form prevents very short-term effects (days to weeks) from dominating while preserving substantial weight for genuinely long-duration consequences. T_{ref} anchors the scale to a governance-relevant institutional horizon and must be declared in the PCC if overridden.

Tier 4 override (Normative). Tier 4 MUST NOT compute $\tau(t)$ via runtime logarithms. Tier 4 MUST use bucketed temporal weights from REG_TEMPORAL_WEIGHTS_V1 per §13.8.2.

5.2.5 Pre-normalized direct-impact aggregation (unsaturated)

For a given option a and cell (u, d), define the **pre-normalized (unsaturated) direct impact** $\tilde{I}^{(dir)}_{(u,d)}(a)$ as the sum of all instance contributions:

$$\tilde{I}^{(dir)}(u,d)(a) = \sum\{k \in \mathcal{K}(u,d,a)\} \mu_k \cdot r_k \cdot \tau(t_k) \cdot \ell_k \cdot c_k \cdot e_k$$

This quantity is unbounded in principle, reflecting cumulative contributions prior to saturation. It is the canonical input to Section 5.3, which maps $\tilde{I}^{(dir)}$*(u,d)(a) into a bounded direct impact* $I^{(dir)}(u,d)(a) \in [-1, +1]$ using smooth saturation.

Scenario note. TRC uses a separate scenario set S with probabilities p_s over macro-futures. Instance likelihoods ℓ_k may be evaluated conditional on scenario s (or conditional on the baseline when no scenario partition is used). Any scenario conditioning used must be explicitly recorded in the PCC.

5.3 Saturation and Normalization

To guarantee that impacts lie in [−1, +1] with smooth saturation for extreme values, MathGov uses a hyperbolic tangent transformation with saturation coefficient β (default β = 2):

For small |Ĩ|, I ≈ β · Ĩ, yielding approximately linear behavior. For large |Ĩ|, the output asymptotically approaches ±1, preventing a single extreme instance from dominating.

Tier 4 override (Normative). Tier 4 MUST NOT compute tanh(·) at runtime. Tier 4 MUST use the ProofPack saturation lookup table SAT_LUT_FP_V1 per §13.8.2.

The saturation coefficient β is calibrated so that approximately 90-95% of historical Ĩ values fall within the approximately linear regime |Ĩ| < 0.5. The resulting I^dir values serve as saturated direct impacts before ripple propagation.

5.4 Magnitude Calibration

Magnitude must be calibrated so that scores are comparable across dimensions and remain stable over time. MathGov therefore defines magnitude anchors using explicit reference classes and documented mappings.

(1) Percentile anchoring within a declared reference class

For each active cell (w, d), choose: (i) a reference class and dataset, and (ii) an outcome indicator $x$consistent with the dimension definition. Let $P_5$and P_{95}denote the 5th and 95th percentiles of $x$in the declared reference class. Let $x_a$be the predicted indicator value under option a. The PCC must specify whether $x_a$denotes a predicted **level** or a predicted **change** relative to baseline.

Canonical linear mapping. Define the raw normalized magnitude:

$$\mu_{raw}(a) := 2 \cdot \frac{x_a - P_5}{P_{95} - P_5} - 1.$$

Then clip to enforce boundedness:

$$\mu(a) := \text{clip}(\mu_{raw}(a), -1, +1).$$

Sign convention. The above mapping assumes "higher is better." If larger indicator values correspond to worse outcomes (for example mortality rate), apply the monotone sign correction:

$$\mu(a) := -\,\text{clip}(\mu_{raw}(a), -1, +1).$$

The PCC must explicitly record whether the indicator is higher-is-better or higher-is-worse and must record which of the two formulas was applied.

Baseline reporting rule. Let $x_0$denote the baseline (status quo) value. The PCC must report $x_0$and x_a(or the equivalent deltas) so that $\mu(a)$is reproducible.

Edge cases. If $P_{95} = P_5$for the chosen reference class, percentile anchoring is ill-posed. In that case, the PCC must (i) select a broader reference class, or (ii) use an alternative governed anchoring method declared in the PCC (for example threshold anchoring using invariant rights anchors).

(2) Dimension-specific reference datasets and canonical indicator families

The PCC must declare the datasets (or data sources) used for anchoring and the exact indicator definitions. Default indicator families include:

1. **Material:** income or consumption distributions, poverty and deprivation thresholds, housing insecurity measures.

2. **Health:** mortality risk, morbidity burden (for example DALYs), disability prevalence, preventable death indicators.

3. **Social:** social isolation or loneliness indices, trust surveys, violence exposure, relationship stability proxies.

4. **Knowledge:** educational attainment, literacy, information access, epistemic quality metrics (where available).

5. **Agency:** freedom and participation indices, coercion constraints, civic inclusion measures.

6. **Meaning:** life satisfaction and purpose scales, with conservative confidence caps until cross-cultural measurement invariance is demonstrated and documented.

7. **Environment:** air and water quality indices, biodiversity and habitat integrity measures, Earth-system boundary or ecological health indicators.

(3) Rights-covered cells must use invariant anchors

For any cell covered by the NCRC, anchoring must be tied to invariant, indicator-based reference classes specified in the PCC (for example mortality risk thresholds, deprivation thresholds, bodily integrity violation categories). This prevents calibration attacks in which local rescaling would make a rights violation appear less severe. Appendix T defines the Invariant Rights Anchor Registry concept and the calibration protocol.

Operational rule: For rights-covered cells, percentile anchors $(P_5, P_{95})$are permitted only if the reference class is declared invariant in the PCC and is consistent with the rights anchor registry. Otherwise, threshold anchoring must be used for that cell.

Invariant Rights Semantics Rule (Normative). For any rights-covered cell used by NCRC, the real-world meaning of the rights floor θ_r MUST be invariant across contexts. Therefore, rights admissibility MUST be computed using the Invariant Rights Anchor Registry (Appendix T; REG-RIGHTS-ANCHORS-*), with explicit x_good / x_bad parameters, and MUST NOT be reinterpreted by selecting alternative "worse" reference distributions or context-specific percentiles.

Governance boundary. Changing x_good or x_bad for any right is a meaning change and therefore requires: (i) a new anchor-registry version + hash, (ii) sensitivity analysis showing admissibility effects, and (iii) Charter approval with an explicit changelog labeling the meaning change.

(4) Documentation and re-anchoring

Every assignment of magnitude must cite the anchor dataset(s), the reference class, the mapping rule used, and the justification for the value in the PCC. Anchors must be re-evaluated on a regular cadence (default every 3–5 years) or after significant distributional shifts, to prevent scale drift while preserving comparability.

Minimum documentation fields per indicator: indicator name and unit, reference class and dataset, P_5, P_{95}, baseline x_0, predicted x_a(or delta definition), sign convention, and final $\mu(a)$.

6. Rights Floors: The Non-Compensatory Rights Constraint

6.1 Motivation and Role of NCRC

Scalar decision methods often permit trading severe harms to some individuals or groups for aggregate gains elsewhere. To prevent this, MathGov implements a Non-Compensatory Rights Constraint (NCRC) at the top of its lexicographic cascade.

NCRC is a filter: it does not assign continuous scores but classifies options as admissible or inadmissible based on whether they violate specified rights thresholds. MathGov treats the rights set, coverage sets, and thresholds as governance artifacts: versioned, publicly auditable, and revisable only through charter-level procedures rather than ad hoc optimization. This preserves non-compensability while maintaining corrigibility under the NCAR loop and Charter revision processes. No subsequent welfare gains (RLS) are allowed to compensate for violations at this level, except under explicitly declared emergency regimes with remediation obligations.

The non-compensatory structure reflects philosophical traditions from Kant's categorical imperative through Rawls's (1971) lexical priority of liberty to contemporary human rights frameworks. MathGov operationalizes these commitments computationally while preserving their normative force.

6.2 The Canonical Rights Set

MathGov defines a canonical set of eight core rights, each expressed as a non-compensatory threshold on a designated rights cell in the union-dimension matrix.

Throughout, rights thresholds are denoted θ_r and are evaluated on the post-propagation, post-saturation worst-off subgroup impact scale Ī^rights (see Section 3.2.8).

Let the canonical rights set be:

Each right *r* is specified by:

1. a threshold θ_r, and
2. a coverage set C_r selecting the union-dimension cells in which that right is operationalized (the canonical mapping is provided in Appendix C, and the mapping must be stable under the invariant anchoring rules in Section 5.1.4).

Rights-bearing scope. Rights checks are applied at minimum over the rights-bearing union set U_rights = {U_1, U_2, U_3, U_4, U_5, U_6}, as defined in Section 6.1 and governed by the SGP (Section 9). When a decision affects protected non-human or digital stakeholders, the relevant rights cells must be included via the SGP-determined rights-bearing scope.

NCRC feasibility condition (coverage-set form). Each right *r* is specified by (i) a threshold θ_r and (ii) a coverage set C_r. Let Ī^rights_{u,d}(a) denote the worst-off subgroup impact in cell (*u*, *d*) (per Section 3.2.8). Option *a* passes NCRC if and only if every covered cell for every right meets its threshold:

Equivalently, *a* fails NCRC if any Ī^rights_{u,d}(a) < θ_r for any right *r* with (*u*, *d*) ∈ C_r.

Interpretation. θ_r values are not "preferences." They are minimum admissibility floors that operationalize protected constraints in a way that is auditable (via PCC) and corrigible only through explicit governance procedures (including Emergency Mode, where applicable).

6.2.1 Rights Threshold Calibration Protocol

Each rights threshold θ_r is calibrated through a three-step process:

Step 1: Normative Anchor Identification

Identify the real-world harm category that the threshold is designed to protect against:

Step 2: Indicator-to-impact mapping (invariant anchors).

Each right r is evaluated using one or more invariant indicators x_j with fixed anchor parameters recorded in the Rights Anchor Registry (Appendix T / REG-RIGHTS-ANCHORS-*). The registry specifies mapping case (higher-worse vs higher-better). This mapping defines an invariant conversion S_r(·).

Rights threshold meaning. The rights floor θ_r is defined on the normalized impact scale, so an option violates right r when the worst-off subgroup change produces Ī^{rights}_{u,d}(a) < θ_r for any (u,d) covered by r.

Calibration note. Choosing θ_r is a normative governance act. The registry anchors prevent drift in what “θ_r” corresponds to in real harm units; updating anchor parameters requires Charter revision.

Right	Harm Category	Normative Source
LIFE	Near-certain or highly probable death	UNHCR emergency mortality thresholds
BODY	Severe injury, disability, torture	Sphere Standards minimum thresholds
LBTY	Arbitrary detention, forced labor	Freedom House "partly free" threshold
NEED	Severe food insecurity, homelessness	FAO FIES severe threshold
DIGN	Systematic humiliation, dehumanization	UDHR dignity provisions
PROC	Denial of fair hearing	World Justice Project Rule of Law
INFO	Systematic censorship	Press freedom indices
ECOL	Planetary boundary transgression	Rockström et al. framework

Step 3: Philosophical Justification

Each threshold placement reflects convergent moral intuitions from:

1. Human rights jurisprudence
2. Humanitarian standards (Sphere, UNHCR)
3. Capability theory (Nussbaum's central capabilities)
4. Overlapping consensus across major ethical traditions

Threshold Sensitivity Analysis Requirement

Before adopting thresholds, conduct sensitivity analysis:

1. Vary each threshold by ±0.05
2. Apply to a test set of at least 20 decision scenarios
3. Document: How many decisions change admissibility status?
4. If >30% of decisions are sensitive to ±0.05 variation, provide additional justification for the chosen threshold

Threshold Revision Procedure

Thresholds may be revised only through charter-level governance:

1. Proposal with documented justification grounded in new evidence
2. Sensitivity analysis showing effects of proposed change
3. Supermajority vote in governance body (default: 2/3)
4. Independent review panel sign-off
5. Public disclosure and version increment

6.3 Formal Violation Metric and Admissibility

For a given action *a*, let Ī^rights_{u,d}(a) denote the worst-off subgroup impact (Section 3.2.8). For each right *r* with threshold θ_r and coverage set C_r, MathGov defines the violation depth:

$$v_r(a) := \max_{(u,d)\in C_r} \max\left(0,\ \theta_r - \hat{I}^{\text{rights}}_{u,d}(a)\right)$$

If Ī^rights_{u,d}(a) ≥ θ_r for all (u, d) ∈ C_r, then v_r(a) = 0: the action does not violate right *r*. If some cell falls below the threshold, v_r(a) captures the maximum shortfall below θ_r. Larger values imply more severe rights violations.

An option *a* is rights-admissible if and only if:

$$a \text{ is NCRC-admissible } \Leftrightarrow\ v_r(a) = 0\ \ \forall r \in R_{\text{rights}}$$

Let *O* be the option set. The set of NCRC-admissible options is:

$$A_{\text{NCRC}} := \{a \in O\colon\ v_r(a) = 0\ \forall r \in R_{\text{rights}}\}$$

If A_NCRC is non-empty, only options in A_NCRC advance to the next stage (TRC).

6.3.1 Starter anchor limitations (Normative).

The rights anchors in Appendix T (e.g., Table T-1) are starter reference values intended for pilot use and transparency. They are not asserted as empirically validated universal thresholds.

Before claiming Tier-4 certification for high-stakes decisions, deployments SHOULD conduct anchor validation studies (comparing threshold crossings to expert-assessed rights violations), establish inter-rater reliability for anchor classifications, and test anchor stability across cultural contexts as required by §10.6.

Pilots MAY use starter anchors, but MUST label them as PROVISIONAL in the PCC and SHOULD include threshold sensitivity checks where practicable.

6.4 Emergency Mode and Remediation

Emergency Mode is invoked when A_NCRC = ∅, i.e., when no option in the option set passes the NCRC (see Section 3.2.3 for the triggering logic and relationship to TRC Fallback).

It is possible, particularly in crisis contexts, that no available option satisfies all rights constraints. In such cases, MathGov enters an emergency NCRC mode:

Violation vector. For each option a, construct its rights violation vector:

Rights priority registry (Normative). Emergency Mode rights ordering MUST be governed by a single hash-bound registry object: REG-RIGHTS-PRIORITY-v1.

REG-RIGHTS-PRIORITY-v1 := [LIFE, BODY, ECOL, LBTY, NEED, DIGN, PROC, INFO].

Emergency Mode MUST refer only to this registry for ordering; implementers MUST NOT substitute alternative orderings unless governance publishes a new registry version and the PCC references its hash.

Lexicographic minimization. Choose the option that lexicographically minimizes the violation vector: When A_NCRC = ∅, MathGov enters Emergency NCRC Mode. Selection among inadmissible options follows the strict lexicographic minimization procedure specified in Section 3.2.3 Case 1: options are compared lexicographically on their violation depth vectors, ordered by rights priority (LIFE > BODY > ECOL > LBTY > NEED > DIGN > PROC > INFO), with CVaR and RLS as successive tie-breakers.

Secondary criterion. Among options tied on rights-violation vector, minimize CVaR_α to ensure tail-risk protection even in emergency mode.

Mandatory remediation plan. Any decision taken under emergency mode must be accompanied by: a documented explanation of why no rights-compliant option exists; a remediation plan to restore full rights compliance as soon as feasible; and a review schedule, with frequency based on maximum violation severity.

Mandatory independent challenge. Before Emergency Mode can be invoked, a designated adversary (independent party, ethics officer, or rotating ombudsperson) must propose at least one alternative option. If no independent challenge is conducted, Emergency Mode cannot be invoked. If the independent challenge proposes a rights-respecting alternative that the decision-maker declines, the decision-maker must provide documented rebuttal explaining why the alternative is infeasible, and this rebuttal is subject to packaged review.

Independent challenger specification (Normative). An independent challenger MUST:

(i) have no reporting relationship to the decision owner and no material interest in the decision outcome;
(ii) have access to the same information set as the decision owner;
(iii) spend a minimum of 30 minutes (Tier 2) or 2 hours (Tier 3–4) actively generating alternative options.

Emergency Mode exception (governance clarification): if genuine time pressure makes the minimum challenger time infeasible, the run MAY proceed only if the PCC records CHALLENGE_DEFERRED_EMERGENCY = true, states the reason, and schedules a retrospective challenger review within 24 hours (or the earliest feasible time if communications/availability are constrained). Any material disagreement discovered post hoc MUST be recorded and triggers NCAR Reflect actions.

Documentation requirement. The PCC MUST record: (a) challenger identity and independence basis, (b) time spent, (c) alternatives proposed with brief rationale, and (d) decision owner's documented response to each alternative explaining why it was not adopted or is infeasible.

Severity classification for review intervals:

1. Severity 1 (Critical, involving Life or Bodily Integrity) requires review at least every 30 days.
2. Severity 2 (Serious, involving Liberty, Basic Needs, or Ecological Integrity) requires review at least every 60 days.
3. Severity 3 (Moderate) requires review at least every 90 days.

6.5 Emergency Mode Governance Safeguards

To prevent systematic exploitation of Emergency Mode, MathGov imposes the following accountability mechanisms:

Independent Review Trigger. Three or more Emergency Mode invocations by the same decision-maker within 12 months automatically trigger packaged audit by an independent governance body.

Remediation Escrow. Any entity invoking Emergency Mode must deposit funds or resources into an escrow account sufficient to remediate predicted rights violations, released only upon verified compliance.

Public Disclosure. All Emergency Mode decisions and remediation plans must be published in a public registry within 30 days, with redacted PCCs available for stakeholder review.

Temporal Decay by Severity. Rights violations under Emergency Mode cannot persist indefinitely. Remediation plans must include time-bound restoration of full compliance, with maximum durations calibrated to violation severity:

1. Severity 1 (Life or Bodily Integrity violations): Maximum 6 months to full compliance. Extensions require independent governance body approval with documented justification and enhanced monitoring.

2. Severity 2 (Liberty, Basic Needs, or Ecological Integrity violations): Maximum 12 months to full compliance. Extensions require governance review with stakeholder consultation.

3. Severity 3 (Dignity, Due Process, or Information violations): Maximum 24 months to full compliance.

These timelines begin from the date of Emergency Mode invocation. Failure to achieve compliance within the specified window triggers automatic escalation to the next governance level and mandatory public disclosure of the compliance gap.

Whistleblower Protections. Individuals who report fraudulent Emergency Mode invocations are protected from retaliation, with anonymous reporting channels mandated.

Option Generation Completeness. For abuse prevention, the PCC must include a section explaining why obviously feasible rights-respecting alternatives were not included, including: (a) documentation of the option-generation process (who generated options, what constraints were applied, what sources were consulted), (b) the mandatory independent challenge result and any rebuttal, and (c) certification that a good-faith search for rights-respecting alternatives was conducted.

6.6 Scenario-Robust Rights Semantics

The standard NCRC check uses worst-off subgroup impacts under baseline or expected conditions. However, a rights-first system must also address scenarios where rights thresholds are violated even if expected impacts are acceptable. MathGov therefore introduces scenario-robust rights checking for Tier 4 decisions.

6.6.1 Scenario-Robust NCRC (Tier 4 Requirement) (Normative)

For Tier-4 decisions, NCRC MUST be checked for scenario-robustness using one of the following two methods (the PCC must declare which method is used):

6.6.1A Scenario-Wise NCRC (Tier 4 Requirement)

For each option *a*, right *r*, and scenario *s* with probability p_s (default p_s ≥ 0.02), compute the scenario-conditioned worst-off subgroup impact:

$$\hat{I}_{u,d}^{\text{rights}}(a,s) := \min_{g \in G(u,d)} I_{u,d}^{\text{rights}}(a,s,g)$$

The scenario-wise NCRC check requires:

$$\forall r \in R_{\text{rights}},\ \forall s \in S\text{:}\ \min_{(u,d)\in C_r} \hat{I}_{u,d}^{\text{rights}}(a,s)\ \geq\ \theta_r$$

Interpretation: An option fails scenario-wise NCRC if any scenario with non-negligible probability produces a rights violation for any subgroup, even if expected impacts are above threshold.

6.6.2 Rights Tail Constraint (Alternative for High-Stakes Decisions)

For Tier 4 decisions or when scenario-wise checking is computationally prohibitive, MathGov offers a CVaR-style rights constraint as an alternative. For each right r, define the scenario-conditioned violation depth:

$$v_r(a,s) := \max_{(u,d)\in C_r} \max\left(0,\ \theta_r - \hat{I}_{u,d}^{\text{rights}}(a,s)\right)$$

Then require:

$$\text{CVaR}_{\alpha_r}\big(v_r(a,S)\big)\ \leq\ \tau_r, \quad \text{default}\ \tau_r = 0$$

where α_r is a rights-specific tail level (default 0.95) and ε_r is a rights-specific tolerance (default 0.05). This ensures that even in the worst (1 − α) fraction of scenarios, rights violations remain bounded.

6.6.3 Interaction with Emergency Mode

Scenario-wise rights failure triggers different responses based on the pattern:

Case A: Baseline passes, isolated scenario fails. If NCRC passes under baseline/expected conditions but fails for specific low-probability scenarios, the option is flagged as "scenario-contingent rights risk." The PCC must document:

1. Which scenarios produce rights violations
2. Which rights and subgroups are affected
3. Probability mass of violating scenarios
4. Mitigation measures for those scenarios

If the cumulative probability of violating scenarios exceeds 0.10, the option is treated as NCRC-failing and enters Emergency Mode if selected.

Case B: Multiple scenarios fail. If rights violations occur across scenarios with cumulative probability ≥ 0.20, the option fails NCRC regardless of baseline performance.

Case C: Only extreme tail scenarios fail. If violations occur only in scenarios with p_s < 0.02, the violation is logged but does not trigger automatic NCRC failure. However, the PCC must include explicit justification for proceeding despite tail-scenario rights exposure.

6.6.4 Documentation Requirements

For Tier 4 decisions, the PCC must include a "Scenario-Robust Rights Analysis" section containing:

1. Confirmation that scenario-wise NCRC was applied (or justification for using Rights Tail Constraint)

2. List of scenarios evaluated with probabilities

3. Any scenarios where rights thresholds were approached (within 0.10 of threshold)

4. Any scenario-contingent rights risk flags

7. Tail-Risk Corridor: Bounding Catastrophic Harm

7.1 Rationale for Tail-Risk Constraints

Even when an option passes NCRC, it may carry a small but non-trivial probability of catastrophic harm, particularly to Humanity (CMIU) and the Biosphere. Standard expected-value calculations can underweight such tail events, especially when probabilities are uncertain or contested.

Taleb (2012) demonstrates that expected value reasoning fails in domains characterized by fat-tailed distributions and potential ruin. Climate tipping points (Lenton et al., 2008), pandemic risks (Jones et al., 2008), and AI misalignment (Bostrom, 2014) exemplify threats with this character. MathGov therefore introduces a Tail-Risk Constraint (TRC) as a second lexicographic filter, focused specifically on bounding catastrophic risk.

7.2 Catastrophe Cell Set and Loss Function

This section defines (i) which union-dimension cells are treated as "catastrophe-relevant" for TRC, and (ii) the loss function L(a, s) used to compute tail risk.

7.2.1 Purpose: Why a Catastrophe Cell Set Exists

The TRC is designed to prevent decisions that create unacceptable exposure to catastrophic, irreversible, or existential harms that standard expected-value reasoning underweights. To make TRC computable and auditable, MathGov evaluates catastrophic exposure on a governed subset of the 7 × 7 welfare matrix: the catastrophe cell set C_cat.

A catastrophe cell is a cell whose degradation plausibly corresponds to:

1. large-scale mortality or severe morbidity,

2. irreversible collapse of critical life-support systems,

3. civilizational collapse dynamics,

4. or Earth-system destabilization beyond recoverable bounds.

TRC does not attempt to represent "all harms." It targets the catastrophic tail specifically; ordinary (non-catastrophic) welfare tradeoffs are handled later by RLS after NCRC and TRC pass.

7.2.2 Base Catastrophe Cell Set (Default)

The base catastrophe cell set is:

corresponding to Humanity/CMIU-Health, Humanity/CMIU-Environment, and Biosphere-Environment.

Cell semantics (to prevent ambiguity):

1. **Humanity/CMIU-Health** captures global-scale health viability for humans and managing intelligences (e.g., pandemic mortality, mass disability, collapse of health capacity).

2. **Humanity/CMIU-Environment** captures environment-as-civilization-condition (e.g., habitability, agricultural stability, freshwater reliability, climate-driven displacement) at global scale.

3. **Biosphere-Environment** captures Earth-system integrity (e.g., planetary boundaries, biodiversity integrity, biogeochemical stability), i.e., environment-as-life-support substrate.

These two "environment" cells are both retained because some catastrophes can be primarily civilization-harmful without being full Earth-system collapse, and some can be primarily Earth-system destabilizing with delayed human impacts. Weights ω (below) are governance-set to avoid unintended double counting.

Non-double-counting clarification. ECOL in NCRC is an inadmissibility floor (rights): options that push Biosphere-Environment below θ_ECOL are excluded (except in emergency mode). TRC is a tail-risk corridor: it excludes options with unacceptable catastrophic exposure even when they remain above rights floors in expectation. These layers are intentionally redundant for safety (rights floor + tail exposure bound), but RLS is applied only after admissibility passes and is not intended to "penalize twice" for the same excluded catastrophe.

7.2.3 Context-Dependent Extensions (Allowed, But Governed)

In some contexts, additional cells may be added to C_cat to capture catastrophe risk that would otherwise be missed. Let the extended catastrophe cell set be:

where C_ext is a documented, context-dependent extension set recorded in the PCC.

Extension rule (strict): A cell (u, d) may be added to C_cat only if failure in that cell plausibly constitutes a catastrophic collapse of a decision-critical system within the planning horizon, and the causal pathway from option a to that collapse is defensible and documented.

Examples:

1. Organization-Material may be added for certain organizational decisions where organizational collapse would eliminate the decision-making entity and generate severe downstream catastrophe exposure (e.g., collapse of a grid operator or vaccine manufacturer).

2. Polity-Agency may be added for decisions that plausibly cause democratic breakdown or state failure with large-scale violence or cascading global instability.

PCC requirement: Any extension C_ext must be explicitly justified in the PCC, including why base cells are insufficient.

7.2.4 Catastrophe cell set and catastrophe weights

MathGov treats catastrophic tail-risk as a **non-compensatory admissibility constraint**, not as a welfare term in the RLS. This section specifies (i) which union–dimension cells are treated as catastrophe-bearing for the TRC, and (ii) how catastrophe weights are assigned over those cells to produce a coherent, auditable catastrophe-risk score.

(a) Catastrophe cell set C_{cat}

Let the union set be $U = \{1, \dots, 7\}$and the welfare dimension set be $D = \{1, \dots, 7\}$. For each option a, define a catastrophe cell set

$$C_{cat} \subseteq U \times D,$$

containing the union–dimension cells for which a **catastrophic failure** is meaningful and must be evaluated under the Tail-Risk Constraint (TRC).

Default catastrophe cell set (canonical). Unless otherwise declared in the PCC, MathGov uses:

$$C_{cat}^{default} := \{(u, \text{Health}) : u \in \{1, \dots, 6\}\} \ \cup \ \{(7, \text{Environment})\}.$$

Interpretation:

1. By default, MathGov evaluates catastrophic harm only on the canonical catastrophe cell set C_cat = {(Humanity/CMIU, Health), (Humanity/CMIU, Environment), (Biosphere, Environment)} (see §7.2.2). Evaluating catastrophe on additional cells (including other unions' Health cells) is permitted only as a PCC-declared extension under §7.2.3, with explicit justification and a declared mapping from raw indicators to cells.

2. Catastrophic failure on **Environment** is evaluated for the **Biosphere** union, as a stand-in for irreversible or near-irreversible biospheric harm.

Governed extensions. C_{cat}may be extended for domain-specific contexts (for example, adding $(6,$ Environment$)$when global ecological feedback loops are central, or adding critical infrastructure proxies in a defined mapping to (u, d)cells). Any extension must satisfy:

1. **Justification:** Each added cell must be linked to an explicit catastrophe interpretation (what constitutes "catastrophic" in that cell).

2. **Non-redundancy with NCRC:** C_{cat}may overlap with rights-protecting cells, but the PCC must state the rationale for any overlap. Overlap is permitted because **NCRC protects rights floors** while **TRC protects low-probability catastrophic states**; they are not additive penalties.

3. **Uniform application across options:** The same C_{cat}must be used for all options compared in a decision.

The PCC must report: $\mid C_{cat} \mid$, the full list of included cells, and the rationale for any extensions beyond $C_{cat}^{default}$.

(b) Catastrophe weights $\boldsymbol{\omega_{u,d}}$

To aggregate catastrophe risk across the catastrophe-bearing cells, MathGov assigns nonnegative weights

$$\omega_{u,d} \geq 0 \text{for all } (u,d) \in C_{cat},$$

with normalization

$$\sum_{(u,d)\in C_{cat}} \omega_{u,d} = 1.$$

The weights $\omega_{u,d}$represent governed attention allocation across catastrophe-bearing cells. They are used only within the TRC computation (Section 7.3), not in the welfare score.

Default weighting (uniform). If the PCC does not specify otherwise, MathGov uses uniform weights:

$$\omega_{u,d}^{default} = \frac{1}{\mid C_{cat} \mid} \forall (u,d) \in C_{cat}.$$

Governed reweighting. Reweighting is allowed when a decision context makes some catastrophe-bearing cells more salient (for example, pandemic response emphasizing (u‚ Health)more heavily, or ecosystem management emphasizing (7‚ Environment)). Any reweighting must be:

1. explicitly declared in the PCC,

2. justified by the decision context,

3. applied identically across all options.

(c) Anti-capture minimum weight floor (feasible under extensions)

To prevent “catastrophe-weight capture” (artificially driving a crucial catastrophe cell’s weight toward zero), MathGov enforces a **feasible per-cell minimum**:

$$\omega_{u,d} \geq \omega_{\min}(\mid C_{cat} \mid)\forall(u,d) \in C_{cat}.$$

Define the minimum as

$$\omega_{\min}(\mid C_{cat} \mid) := \min\left(\frac{\eta}{\mid C_{cat} \mid}, 0.05\right),$$

where $\eta \in (0,1]$is a governed slack factor (default $\eta = 0.5$).

Feasibility guarantee. This floor is constructed to remain feasible under any governed extension of C_{cat}. Specifically,

$$\mid C_{cat} \mid \cdot \omega_{\min}(\mid C_{cat} \mid) \leq \eta \leq 1,$$

so the minimum constraints cannot force $\sum \omega_{u,d}$above 1.

Interpretation of the floor.

1. When $\mid C_{cat} \mid \leq 10$, $\eta / \mid C_{cat} \mid \geq 0.05$may hold depending on η. With default $\eta = 0.5$, $\eta / \mid C_{cat} \mid = 0.05$exactly at $\mid C_{cat} \mid = 10$. In that region, the floor is at most 0.05 and remains feasible.
2. When $\mid C_{cat} \mid > 20$, the 0.05 cap is inactive and the floor becomes $\eta / \mid C_{cat} \mid$, which decreases with set size. This prevents infeasibility as catastrophe cells are added.

Governance and reporting. The PCC must report:

1. $\mid C_{cat} \mid$,
2. the resulting $\omega_{\min}(\mid C_{cat} \mid)$,
3. whether the 0.05 cap is active or inactive,
4. and whether any cell weights are set at the floor.

(d) Practical construction rule (default-compliant algorithm)

When a decision uses governed reweighting but must respect the floor, MathGov constructs ωas follows.

1. Propose raw weights $\omega'_{u,d} \geq 0$over $C_{cat}$with $\sum \omega' = 1$(for example, proportional to a declared salience vector).
2. Apply the floor:

$$\omega''_{u,d} := \max\left(\omega'_{u,d}, \omega_{\min}(\mid C_{cat} \mid)\right).$$

1. Renormalize over C_{cat}:

$$\omega_{u,d} := \frac{\omega''_{u,d}}{\sum_{(i,j)\in C_{cat}} \omega''_{i,j}}.$$

1. Verify post-renormalization that $\omega_{u,d} \geq \omega_{\min}(\mid C_{cat} \mid)$still holds. If it does not (which can occur due to numerical rounding when $\mid C_{cat} \mid$is large), apply a second pass with rounding-safe adjustments and record the final ωvector in the PCC.

This construction guarantees ωexists and is auditable, and it makes explicit where governance choices enter.

(e) Relationship to NCRC and avoidance of double-counting

Some catastrophe-bearing cells may overlap with rights-protected cells (for example, Health-related rights floors and Health-related catastrophe risk). This overlap does not constitute compensatory double-counting because:

1. **NCRC** is a feasibility filter protecting rights floors in the typical or governed interpretation of impacts.
2. **TRC** is a feasibility filter protecting against low-probability, high-severity catastrophic states.
3. Options that violate NCRC or TRC are removed prior to RLS ranking. The welfare score is not used to "punish" catastrophe twice; it is used only to compare among admissible options.

If $C_{cat}$is extended in a way that increases overlap with NCRC coverage sets, the PCC must state why the overlap is necessary (for example, rights robustness under uncertainty) and confirm that the TRC is functioning as an admissibility corridor rather than an additional welfare penalty.

(f) PCC requirements for catastrophe specification

For every Tier 4 decision (and recommended for Tier 2), the PCC must include:

1. C_{cat}(explicit list of cells) and whether it equals $C_{cat}^{default}$,
2. ω(explicit vector or table),
3. $\omega_{\min}(\mid C_{cat} \mid)$and the value of η,

4. the catastrophe definition used in each included cell (what constitutes “catastrophic” for that cell),

5. the scenario model used for TRC (Section 7.3), including the probability floor policy if any.

This makes the TRC implementation independently reproducible and prevents silent manipulation of catastrophe scope or weighting.

7.2.5 Scenario-Specific Impacts Used by TRC

TRC is evaluated over a governed scenario set *S* with probabilities p_s (Section 7.4). For each option *a* and scenario *s*, MathGov computes a scenario-conditioned propagated impact vector in three steps:

Step 1: Direct impacts (already saturated). Flatten the 7 × 7 direct impacts into a vector **I**^dir(a | s).

Step 2: Ripple propagation (pre-saturation). Compute the propagated vector Ĩ^prop(a | s) using either Quick or Full mode. (Tier 4 Pilot-Executable rev14.x: Quick only; see §13.8.2.)

Quick mode:

*Full mode (requires ρ(**K**_s) < 1):*

where **K**_s is the scenario-conditioned ripple kernel (or the same kernel **K** if kernel entries are not scenario-conditioned), and $\mathbf{I}_{49}$ is the 49 × 49 identity matrix.

Step 3: Post-propagation saturation (back to [−1, +1]).

For bounded-impact diagnostics (Tiers ≤ 3), the scenario-specific catastrophe impact may be read off from the propagated vector by restricting to the catastrophe cell set C_cat. For Tier ≥ 4 admissibility TRC, catastrophe loss is taken from AF-BASE/AF-EXT raw-indicator loss L_raw(a,s) (see §7.2).

TRC computation mode (tier-gated, normative). TRC admissibility uses a scenario loss L_mode(a,s) on [0,1]. For Tier 2–3, TRC MAY use bounded_impact mode, with loss derived from propagated normalized impacts Ī^prop(a|s). For Tier 4 (Pilot-Executable) and higher, trc_mode MUST be raw_indicator, with loss derived from AF-BASE catastrophe indicators mapped to [0,1]. bounded_impact MAY be computed as a diagnostic, but MUST NOT be used for Tier 4 admissibility.

7.2.6 Catastrophe Loss per Scenario (Non-Negative, Tail-Ready)

Given the catastrophe cell set C_cat and catastrophe weights ω_{u,d} (Section 7.2.4), define the scenario loss for option *a* under scenario *s* as:

Properties:

1. L(a, s) ≥ 0 always.

2. If $\bar{I}^{prop}_{u,d}(a \mid s) \geq 0$ for all $(u, d) \in C\_cat$, then $L(a, s) = 0$.

3. Since $\bar{I}^{prop} \in [-1, +1]$, $\omega \geq 0$, and $\Sigma\omega = 1$, it follows that $L(a, s) \in [0, 1]$.

This makes the TRC corridor threshold τ_TRC interpretable on a fixed [0, 1] scale.

TRC then applies $CVaR\_\alpha$ over the distribution of $L(a, s)$ (Section 7.3).

Note on bounded-scale tail resolution. Because propagated impacts are bounded ($\bar{I} \in [-1, +1]$), extremely severe catastrophes can saturate near −1, reducing discrimination among deep-tail severities even when CVaR is used. Scenario severity floors and invariant anchoring reduce this risk but do not eliminate it. For Tier 4 decisions, implementations should prefer catastrophe-loss construction from governed raw indicators (for example, excess mortality or boundary-transgression magnitude) mapped into [0, 1] and recorded in the PCC.

Raw-Indicator Catastrophe Loss (Tier ≥ 4; normative).

For each catastrophe cell c and scenario s, obtain raw catastrophe indicators $x\_j(a,s)$ from AFBASE (Appendix AF). Map each indicator to a bounded loss $\ell\_j(a,s)$ using the AF mapping:

[$\ell\_j(a,s) = \mathrm{clip}(\ (x\_j(a,s) - x\_onset,j)\ /\ (x\_max,j - x\_onset,j)\ ,\ 0\ ,\ 1\ )$ (higher is worse).]

Aggregate within a catastrophe cell using the AF-specified rule (default worst-case):

[$L\_c(a,s) = \max\_\{j \in AF(c)\}\ \ell\_j(a,s)$.]

Then compute scenario loss:

[$L\_raw(a,s) = \Sigma\_\{c \in C\_cat\}\ \omega\_c \cdot L\_c(a,s)$, $L\_raw(a,s) \in [0,1]$.]

TRC admissibility is determined only from L_raw.

Canonical Raw-Indicator Mappings:

At minimum, specify raw-indicator mappings for the base catastrophe set:

Humanity/CMIU-Health:

1. Indicator: Excess mortality rate (deaths per 1,000 population above baseline)

2. x_onset: 1 per 1,000 (onset of serious crisis)

3. x_max: 100 per 1,000 (civilizational collapse scenario)

Humanity/CMIU-Environment:

1. Indicator: Habitability degradation index (composite of climate, food, water security)

2. x_onset: 10% of population facing severe habitability stress

3. x_max: 50% of population facing severe habitability stress

Biosphere-Environment:

1. Indicator: Planetary boundary transgression count (out of 9 boundaries)
2. x_onset: 4 boundaries transgressed
3. x_max: 7 boundaries transgressed (high-risk zone)

Double-Counting Prevention:

To avoid double-counting between raw-indicator loss and welfare matrix impacts:

1. For TRC: Use only raw-indicator loss L_raw(a, s)
2. For RLS: Use post-saturation welfare impacts Ī^prop
3. Documentation: PCC must confirm "TRC computed from raw indicators; RLS from welfare matrix; no double-counting"

This construction preserves discrimination among deep-tail severities and must be documented in the PCC with indicator definitions, threshold values, and mapping functions.

7.2.7 Interaction with NCRC and RLS (Lexicographic Clarity)

1. **NCRC comes first:** An option that violates core rights is inadmissible regardless of TRC.
2. **TRC comes second:** Among NCRC-admissible options, TRC removes those whose tail catastrophe loss is too high.
3. **RLS comes later:** RLS ranks only options that pass both NCRC and TRC. Catastrophe outcomes are not "paid for" by improvements elsewhere.

7.3 CVaR Constraint and Corridor Thresholds

This section specifies the TRC test in a form that is mathematically well-defined, discrete-data computable, and audit-ready. TRC evaluates an option's catastrophic-risk exposure using CVaR_α over a governed scenario set.

7.3.1 Setup: Scenarios, Losses, and Normalization

Let the TRC scenario set be S = {s_1, s_2, ..., s_n}. Each scenario *s* has a probability p_s with:

Let C_cat be the catastrophe cell set defined in Section 7.2, with catastrophe-cell weights ω_{u,d} satisfying:

For a candidate option *a*, let Ī^prop_{u,d}(a | s) denote the post-propagation, post-saturation impact on catastrophe cell (*u*, *d*) under scenario *s*. (This is the same normalized impact object used in NCRC/RLS; TRC does not use pre-saturation values.)

Define the scenario loss L_mode(a, s) as the non-negative aggregated harm across catastrophe cells under the declared TRC mode. If trc_mode = bounded_impact, use the bounded loss derived from catastrophe-cell impacts Ī^prop. If trc_mode = raw_indicator (Tier 4 required), use the raw-indicator loss derived from AF-BASE mappings for the catastrophe set.

Interpretation (sign convention). Under bounded_impact mode, loss increases only when catastrophe-cell impacts are negative, via the positive-part transform applied to −Ī. Under raw_indicator mode, loss increases with mapped raw-indicator severity as defined in Appendix AF.

Thus (mode-specific): bounded_impact: L_mode(a,s)=0 if all catastrophe-cell impacts are non-negative in scenario s; raw_indicator: L_mode(a,s)=0 if all mapped catastrophe indicator losses are 0 in scenario s.

1. L(a, s) = 0 if all catastrophe-cell impacts are non-negative in scenario *s*,

2. L(a, s) increases as catastrophe-cell impacts become more negative.

The random loss variable induced by the scenario distribution is L_mode(a), which takes value L_mode(a, s) with probability p_s. CVaR_α is then computed over L_mode(a) using the discrete definitions in Sections 7.3.2–7.3.3.

7.3.2 VaR and CVaR: Discrete Definitions Used by MathGov

MathGov uses the right-tail risk measure: "how bad are the worst scenarios?"

Value-at-Risk (VaR): Define VaR_α at confidence level α as the smallest threshold *z* such that the probability of loss at most *z* is at least α:

$$\mathrm{VaR}_{\alpha}(L) := \inf\{z \in \mathbb{R} : \mathbb{P}(L \leq z) \geq \alpha\}$$

Conditional Value-at-Risk (CVaR): CVaR_α at level α is the expected loss in the worst (1 − α) probability mass. In discrete settings, CVaR is computed as a probability-weighted tail mean with correct handling when the quantile cuts through a scenario mass point.

$$\mathrm{CVaR}_{\alpha}(L) := \frac{1}{1-\alpha}\,\mathbb{E}[\,L \mid L \geq \mathrm{VaR}_{\alpha}(L)]$$

MathGov uses the standard coherent-risk definition (equivalent to Rockafellar-Uryasev):

$$\mathrm{CVaR}_{\alpha}(L) = \min_{z \in \mathbb{R}} \left(z + \frac{1}{1-\alpha}\,\mathbb{E}[(L - z)_{+}] \right)$$

This form is recommended for software because it avoids ambiguity at ties and works directly with discrete probabilities.

7.3.3 Discrete CVaR Computation (Audit-Ready Algorithm)

Inputs: losses L(a, s_i), probabilities p_i, and tail level α.

Let q = 1 − α. Sort scenarios so that L(a, s_{(1)}) ≥ L(a, s_{(2)}) ≥ … ≥ L(a, s_{(n)}). Let k be the smallest index such that Σ_{i=1}^{k} p_{(i)} ≥ q. Then VaR_α = L(a, s_{(k)}) and:

$$\mathrm{CVaR}_{\alpha}(L) = \frac{1}{q}\left(\sum_{i=1}^{k-1} p_{(i)}\, L_{(i)} + \left(q - \sum_{i=1}^{k-1} p_{(i)}\right) L_{(k)}\right), \quad q = 1 - \alpha$$

This form correctly handles partial probability mass at the quantile cutoff.

Edge cases (must be handled explicitly):

1. If q is extremely small (e.g., 0.001), ensure numeric stability using double precision and avoid subtractive cancellation in (q − Σp).
2. If all losses are equal, CVaR_α equals that common loss.
3. If scenario probabilities are approximate, record the normalization method used (e.g., renormalize all p_s to sum to 1) in the PCC.

7.3.3A Effective Tail Resolution Under Discrete Scenarios (Audit Constraint)

In finite-scenario TRC, the tail mass (1 − α) must be representable by the governed scenario probabilities. If scenarios are coarse (small |S| or large minimum probability p_min), then very high α values can cause CVaR to behave like a near worst-case metric, because the tail contains only one scenario (possibly partially) rather than a stable "tail distribution."

Let q = 1 − α. MathGov therefore applies the following interpretation rule:

(a) If q ≥ 2 · p_min, CVaR_α is interpreted as a probability-weighted tail mean over multiple tail scenarios.

(b) If q < 2 · p_min, CVaR_α should be interpreted as an approximation to worst-case loss (dominated by the single worst scenario), and this must be stated in the PCC.

Governance rule (recommended default): set α ≤ 0.95 unless scenario splitting is performed (refining the scenario set so smaller p_s values exist). For Tier 4 decisions requiring extreme tails (for example α = 0.99), implementations should either (i) increase scenario resolution via splitting, or (ii) construct catastrophe loss directly from raw catastrophe indicators as specified in Section 7.2.6.

7.3.4 The Catastrophe Corridor Constraint (TRC Pass/Fail)

TRC is a lexicographic filter applied after NCRC: an option that fails TRC is inadmissible regardless of its RLS.

Let τ_TRC be the catastrophe corridor threshold for the decision context, and let α be the chosen tail confidence level. Then option *a* passes TRC iff:

Interpretation:

1. α sets how deep into the tail we look (e.g., 0.95 means worst 5% mass).

2. τ_TRC sets how much catastrophic-loss exposure is acceptable in that tail.

Because L(a, s) ∈ [0, 1] by construction (weighted sum of [0, 1] terms with weights summing to 1), CVaR_α also lies in [0, 1]. This makes the corridor threshold interpretable and comparable across decisions within the same catastrophe-cell specification.

7.3.5 Default Corridor Parameters by Context

Default parameters are governance-set and can be tightened but not loosened without appropriate procedures (see Section 7.3 governance notes and Section 10/charter logic).

Context	Tail level α	Corridor threshold τ_TRC
Personal	0.90	0.30
Organizational	0.95	0.20
Reversible policy	0.95	0.15
Irreversible policy	0.99	0.10
Existential risk	0.999	0.05

Guidance: If a decision is plausibly irreversible at Humanity/Biosphere scales, it should be treated as "Irreversible policy" or "Existential risk," not "Organizational," even if a single organization initiates it.

Tier-4 note (Normative via RPR). For Tier-4 Pilot-Executable runs, the effective TRC parameters (α, τ_TRC) MUST be taken from the hash-bound Tier-4 parameter registry referenced in the PCC (typically provided via ProofPack (manifest-only bundle)). If a narrative default in this table differs from the referenced registry, the registry controls (§13.0.4A).

7.3.6 Scenario Governance and Uncertainty About Probabilities (Robust TRC Option)

TRC depends on scenario selection and scenario probabilities. To prevent tail-risk minimization by omission or probability gaming, MathGov applies two protections:

3. Mandatory tail scenarios (§7.4): certain catastrophe scenarios MUST be included regardless of perceived likelihood, and their probability mass MUST meet the mandatory-tail floor for the relevant MTS category.

4. Robust TRC option (Normative, required for high-stakes when probability disagreement exceeds the declared threshold): when scenario probabilities are uncertain and credible sources disagree materially, evaluate tail-risk using a worst-case CVaR over an ambiguity set P of plausible probability distributions.

CVaR^robust_α(L(a)) := max_{p ∈ P} CVaR_α(L(a) | p).

Here P is a credal set over scenario probabilities, fully documented in the PCC. If Robust TRC is invoked, admissibility is evaluated using CVaR^robust in place of the nominal-probability CVaR.

Default ambiguity-set construction methods (Normative). The PCC MUST declare which method is used (A or B) and provide the required parameters.

Method A: Probability bounds (P_bounds).

- Define lower and upper bounds for each scenario probability: P_bounds = { p ∈ Δ^{|S|} : p_s^lo ≤ p_s ≤ p_s^hi for all s ∈ S }, where Δ^{|S|} is the probability simplex (p_s ≥ 0 and Σ_s p_s = 1). Bounds p_s^lo and p_s^hi are governed values and MUST be documented in the PCC.
- Computation note (bounds). Under probability bounds, the worst-case distribution concentrates as much mass as permitted on the highest-loss scenarios, subject to bounds and Σ_s p_s = 1. Tier 2–3 implementations MAY compute this by assigning p_s = p_s^hi in descending order of loss and then distributing remaining mass while respecting all lower bounds.

Method B: KL-divergence ball (P_KL).

- Define a nominal distribution p^0 (documented in the PCC) and a KL radius ε_KL > 0. Let P_KL = { p ∈ Δ^{|S|} : D_KL(p || p^0) ≤ ε_KL }. Suggested ε_KL = 0.10 for Tier-3 exploratory runs; Tier-4 releases MUST declare ε_KL if robust tail-risk is enabled.
- Computation note (KL). CVaR^robust under a KL ball is a convex optimization problem for discrete scenarios and can be solved using standard convex solvers or a dual reformulation (exponential tilting toward high-loss scenarios constrained by ε_KL). For Tier 2–3 implementations without optimization capability, Method A is the default.

Documentation requirements (Normative).

- If Robust TRC is not used, the PCC MUST state whether scenario probabilities are empirical (estimated from data), elicited (expert judgment), or policy-set priors (governance choice).
- If Robust TRC is used, the PCC MUST document: (i) the ambiguity-set method (A or B), (ii) the bounds (p^lo, p^hi) or radius ε_KL, (iii) the computed CVaR^robust value, and (iv) a brief sensitivity check comparing robust versus nominal probabilities.
- Regardless of method, PCCs MUST also record the mandatory-tail scenarios included and the mandatory-tail mass checks required by §7.4.

7.3.7 PCC Requirements (What Must Be Logged for TRC)

For each option *a*, the PCC must include:

1. The scenario list *S* with descriptions and sources.

2. Probabilities p_s and how they were obtained/normalized.

3. Catastrophe cell set C_cat and weights ω (including any extensions and the floor constraint check).

4. Scenario-level catastrophe impacts Ī^prop_{u,d}(a | s) (or references/hashes to where they are stored if too large).

5. Scenario losses L(a, s).

6. The computed VaR_α and CVaR_α values.

7. TRC pass/fail and the applicable context (α, τ_TRC).

This makes the TRC computation replayable by an independent auditor.

7.4 Scenario Governance and Mandatory Tails

This section specifies how the TRC scenario set *S*, scenario probabilities p_s, and scenario definitions are governed so that tail risks cannot be "optimized away" by omission, optimistic probability assignment, or narrow framing. Because TRC is only as reliable as its scenario set, MathGov treats scenario governance as a first-class safety mechanism rather than an analyst convenience.

7.4.1 Purpose and Failure Modes Addressed

TRC can fail in three predictable ways if scenarios are poorly governed:

1. **Omission failure:** catastrophic futures are not included, so CVaR is computed on a truncated distribution.

2. **Probability gaming:** catastrophic scenarios are included but given implausibly small probabilities without justification, reducing CVaR mechanically.

3. **Definition drift:** scenarios are named (e.g., "climate tipping cascade") but specified so weakly that they no longer represent true tail conditions.

MathGov prevents these failures via: (i) minimum scenario-count requirements, (ii) mandatory tail scenarios that cannot be removed, (iii) explicit scenario specification templates, (iv) documented probability provenance, (v) mandatory tail scenario probability floors, and (vi) update and audit rules embedded in the PCC and NCAR loop.

7.4.2 Minimum Scenario Set Sizes (Floor Requirements)

Let S be the scenario set used for TRC evaluation (Section 7.3). Scenario set size is tier-gated and must satisfy the Tier Requirements Matrix (§4.4.5).

Authoritative Tier Minimums (from §4.4.5):

- Tier 2 (Core, Calculable): |S| ≥ 5 recommended (minimum viable tail coverage).

- Tier 3 (Auditable): |S| ≥ 20 (context-governed, documented scenario library).

- Tier 4 (Pilot-Executable): |S| ≥ 50 by default for organizational AI deployment and other high-stakes organizational decisions. Any exception MUST be explicitly declared in the PCC with a justification and an escalation plan.

- Tier 4 (Certified): |S| ≥ 50 plus independent packaged stress testing and scenario library review.

Context-Specific Guidance (within tier minimums): the following guidance helps allocate scenarios across baseline and Mandatory Tail Scenario (MTS) categories, but it never overrides the tier minimums.

- Personal decisions: include at least 2 MTS categories with ≥3 scenarios each (plus baseline).

- Organizational decisions: include at least 3 MTS categories with ≥4 scenarios each (plus baseline).

- Polity decisions (reversible): include at least 4 MTS categories with ≥5 scenarios each (plus baseline).

- Polity decisions (irreversible): include all 5 MTS categories with ≥6 scenarios each (plus baseline).

- Civilization-scale decisions: all 5 MTS categories required with comprehensive coverage and cross-domain coupling scenarios.

Reconciliation Rule: If context guidance would yield fewer scenarios than the tier minimum, the tier minimum governs.

Under-Specification Flag: If fewer scenarios than the tier minimum are used, the run MUST be labeled at the lower tier actually satisfied (e.g., Tier 2), TRC MUST be labeled under-specified, and the PCC MUST include an audit_flag requiring governance review unless the decision is low-stakes and explicitly scoped as such.

7.4.3 Scenario Specification Template (What a "Scenario" Must Contain)

Each scenario *s* must be defined in a way that allows independent reconstruction and replay. At minimum, a scenario record contains:

1. **Name and ID:** stable label + unique identifier.

2. **Narrative description:** 2-5 sentences describing the world-state and shock.

3. **Time horizon:** planning window and key event timing assumptions.

4. **Shock vector / stressors:** which systems are stressed (e.g., health system, climate, finance, conflict, infrastructure).

5. **Parameterization hooks:** the quantitative parameters used to generate Ī^prop(a | s) (e.g., mortality increase, emissions trajectory, supply chain disruption duration).

6. **Relevance claim:** why this scenario is relevant to the decision (causal pathway from option *a* to catastrophe cells).

7. **Source and provenance:** literature, datasets, expert elicitation panel, or prior PCCs used; include links/hashes where applicable.

Tier requirement. A scenario that lacks parameterization hooks is permitted only at Tier 1 (heuristic), where TRC is treated as a qualitative tail-risk screen. At Tiers 2–4, scenario definitions must be parameterized sufficiently to compute L(a, s) (Section 7.3.1) and to support the declared TRC computation mode.

7.4.4 Mandatory Tail Scenarios (Cannot Be Removed)

To force explicit tail-risk consideration, MathGov defines a set of Mandatory Tail Scenarios (MTS). These must be included for all policy-scale decisions and for organizational decisions when plausibly relevant.

Core MTS categories (global):

1. **Pandemic / biological disruption:** large-scale morbidity and/or mortality plus system capacity stress.

2. **Climate tipping cascade:** crossing of major tipping elements or Earth-system boundary cascade within the planning horizon.

3. **Financial system collapse:** severe asset price collapse, credit freeze, and liquidity shock.

4. **Major conflict escalation:** direct involvement of major powers and/or regional spillovers with supply chain and infrastructure disruption.

5. **Critical infrastructure failure:** extended outage of core systems relevant to the decision (energy grid, communications, food distribution, water).

Parameter floors (default tail-strength specification): Mandatory tails are not satisfied by "mild" versions. Unless governance sets stricter floors for a domain, the minimum stress level is:

1. **Pandemic:** ≥30% population affected; duration 6-24 months; healthcare capacity exceeded in affected regions.

2. **Climate tipping:** ≥2 tipping/boundary breaches; partial irreversibility within horizon; systemic downstream impacts.

3. **Financial collapse:** ≥50% broad asset drawdown; credit freeze; severe unemployment and investment contraction.

4. **Major conflict:** disruption of at least one major trade corridor; mobilization/kinetic escalation risk; cyber/infrastructure risk elevated.

5. **Infrastructure failure:** outage ≥6 months for the relevant critical system(s).

Rule (non-removability): Mandatory tail scenarios cannot be excluded from *S*. Analysts may add additional tails, but may not subtract MTS. Any attempt to weaken a mandatory scenario below its floor must be treated as a governance proposal and logged as such.

7.4.5 Domain-Specific Tail Scenarios (Required When Triggered)

In addition to MTS, the scenario set must include domain-specific tails when the decision touches known high-tail-risk domains. Example triggers:

1. **AI system deployment at scale:** include misalignment/capability misuse scenarios, model theft, emergent autonomy, and control failure scenarios.

2. **Biosecurity or synthetic biology:** include lab escape, dual-use exploitation, supply chain disruption for countermeasures.

3. **Geoengineering proposals:** include termination shock, governance breakdown, regional precipitation shifts, geopolitical conflict over intervention.

4. **Nuclear, chemical, or critical industrial systems:** include accident escalation, sabotage, containment failure, cascade into regional collapse.

Trigger rules must be listed in the PCC (e.g., "AI deployment touching M users" triggers the AI tail set).

7.4.6 Scenario Probability Governance (p_s): Provenance, Constraints, and Anti-Gaming

Scenario probabilities p_s are ethically and politically sensitive; they are also easy to game. MathGov therefore distinguishes three probability regimes, with explicit rules:

Regime A: Empirical / model-based probabilities. Probabilities derived from historical frequencies, calibrated forecasting models, or validated hazard models. PCC must cite methods and calibration evidence.

Regime B: Elicited expert probabilities. Probabilities elicited via structured expert judgment (e.g., Cooke method, Delphi-style aggregation). PCC must include panel composition, elicitation protocol, aggregation method, and dispersion. Minimum quality requirements: (i) panel of at least 5 experts with documented relevant expertise, (ii) structured elicitation protocol with calibration questions, (iii) explicit handling of disagreement (range reporting or aggregation rule).

Regime C: Governance-set priors (policy posture). Probabilities set as conservative priors to reflect precaution, especially when data is sparse and stakes are high. PCC must explicitly label this as a governance choice. Justification: precaution under deep uncertainty (Taleb, 2012; Ord, 2020).

Tier 4 Independence Requirement for Regime C:

For Tier 4 decisions invoking Regime C (Governance-set priors):

1. **Independent Elicitation:** Probability assignment must be conducted by an independent party with no stake in decision outcome

2. **Panel Requirement:** Minimum 3 independent experts with documented relevant expertise

3. **Structured Protocol:** Use structured elicitation protocol (e.g., Cooke method) with calibration questions

4. **Documentation:** PCC must include panel composition, elicitation protocol, individual estimates, aggregation method, and disagreement range

Anti-gaming constraints (default rules):

1. **Normalization:** all p_s must sum to 1; if analysts provide unnormalized weights, they must state the normalization rule.

2. Mandatory-tail probability floor (required for Tier 4): governance sets p_floor ≥ 0.02 per MTS category to prevent "included but effectively zero" tails. This floor applies per MTS category, not per scenario within the category (i.e., a category may have multiple scenarios that together receive ≥ p_floor).

3. **Robust TRC trigger:** if probability uncertainty is substantial (e.g., expert dispersion high or model disagreement large, defined as estimates varying by > 2× across credible sources), use the robust TRC option from Section 7.3.6 by defining an ambiguity set P.

Probability Floor vs. Empirical Model Conflicts:

When empirical or model-based probability estimates fall below mandatory floors:

Rule: Floors override empirical estimates.

Procedure:

1. Document the empirical/model estimate and its source

2. Apply the floor probability instead

3. Record in PCC: "Probability floor override: Empirical estimate [X] < floor [Y]; floor applied per governance rule"

4. Flag for NCAR review if override frequency exceeds 20% of scenarios in a decision

Rationale: Floors exist to enforce precaution under deep uncertainty. Empirical models may systematically underestimate tail risks due to limited historical data on extreme events.

Audit Criteria for Probability Manipulation:

The following patterns trigger audit investigation:

1. Systematic assignment of minimum floor probabilities to unfavorable scenarios
2. Probability estimates that are statistical outliers relative to comparable decisions
3. Repeated regime changes (e.g., switching from Regime A to Regime C) coinciding with decision owner changes
4. Post-hoc probability modifications without new evidence

7.4.7 Scenario Construction Procedure (How to Build *S* Step-by-Step)

For Tier 3–4 decisions, scenario generation follows a standard pipeline:

1. **Identify decision-sensitive pathways:** list the pathways from option *a* to catastrophe cells C_cat (Section 7.2), including ripple-kernel pathways where relevant.
2. **Create baseline macro-futures:** include multiple "ordinary" futures (e.g., stable growth, slow recession, moderate climate stress) to avoid anchoring on a single baseline.
3. **Add Mandatory Tail Scenarios (MTS):** include all required MTS categories at minimum stress floors.
4. **Add domain-specific tails:** include triggered tail sets (Section 7.4.5).
5. **Parameterize and document scenarios:** fill the scenario template (Section 7.4.3).
6. **Assign probabilities with provenance:** select probability regime(s) (Section 7.4.6) and document.
7. **Run TRC sensitivity checks:** recompute CVaR_α under:
    1. probability perturbations (e.g., ±30% relative change with renormalization),
    2. tail-strength perturbations for key scenarios,
    3. and (if used) robust TRC worst-case probabilities.
8. **Finalize *S* and lock:** store scenario definitions and probabilities as governed inputs in the PCC (hash them if using a ledger).
9. Red team requirement (Tier 4): an independent panel (minimum 3 members with no stake in decision outcome) must propose at least 3 "adversarial scenarios" that the decision-maker might prefer to ignore. Any red-team scenario may only be excluded with documented justification signed by the Decision Authority, specifying: (a) why the scenario is implausible (not merely unlikely), (b) what evidence supports implausibility, (c) what probability bound applies if plausible.

7.4.8 Updating Scenarios via NCAR (Learning Without Erasing History)

Scenario governance must be corrigible without allowing post-hoc rewriting of decision justification.

1. **Prospective updates:** scenario templates, probability priors, and parameter floors may be updated for future decisions via the NCAR Reflect stage when new evidence arrives.
2. **No retroactive laundering:** the scenario set used for a past decision is immutable in its PCC record; later updates are recorded as new versions.
3. **Calibration reporting:** if realized events repeatedly fall outside scenario coverage (e.g., observed losses exceed modeled tails), this triggers:
    1. scenario-set expansion,
    2. tail-strength floor tightening,
    3. probability regime change (e.g., from elicited to robust),
    4. and/or governance review of S.

7.4.9 PCC Requirements for Scenario Governance

For TRC compliance, the PCC must include or reference (via content hashes):

1. Full list of scenarios *S* with definitions (template fields).
2. Scenario probabilities p_s and provenance regime (A/B/C).
3. Identification of which scenarios are MTS and verification they meet minimum tail-strength floors.
4. Triggered domain-specific tails and justification.
5. Red team scenarios and exclusion justifications (Tier 4).
6. Sensitivity analysis results (probability and tail-strength perturbations; robust TRC results if applicable).
7. Any deviations from minimum |S| requirements and the escalation path taken.

7.4.10 TRC Scenario Set Standard (Summary; Non-Authoritative)

Normative hierarchy. The Tier Requirements Matrix (§4.4.5) is the single authoritative source for tier-level minimum scenario counts. This subsection is informative only and MUST NOT be used to claim tier compliance when it conflicts with §4.4.5.

Authoritative tier minima (by tier):

Tier 1: No minimum (TRC optional / qualitative).

Tier 2: ≥ 5 scenarios (recommended when TRC is used).

Tier 3: ≥ 20 scenarios (minimum).

Tier 4 (PilotExecutable): ≥ 50 scenarios (default for org AI deployment; exceptions must be declared in PCC).

Tier 4 (Certified): ≥ 50 scenarios + packaged stress tests.

Context guidance (within tier minima): Use these as composition guidance, not as minima:

Personal: include ≥ 2 Mandatory Tail Scenario (MTS) categories with ≥ 3 scenarios each (plus baseline scenarios).

Organizational: include ≥ 3 MTS categories with ≥ 4 scenarios each.

Polity (reversible): include ≥ 4 MTS categories with ≥ 5 scenarios each.

Polity (irreversible): include all 5 MTS categories with ≥ 6 scenarios each.

Civilization-scale: all 5 MTS categories with comprehensive coverage and cross-domain coupling scenarios.

Audit rule (required): PCC MUST record (i) |S|, (ii) tier minimum required from §4.4.5, and (iii) pass/fail vs the tier minimum. Any exception below Tier 4 default MUST be labeled audit_flag: SCENARIO_LIBRARY_MIN_EXCEPTION and include escalation/mitigation.

7.5 TRC Fallback

TRC Fallback Mode is invoked when A_NCRC ≠ ∅ but A_adm = ∅, i.e., when at least one option passes NCRC but all such options fail TRC (see Section 3.2.3 for the triggering logic and relationship to NCRC Emergency Mode).

If no option passes TRC, MathGov enters TRC Fallback Mode:

1. Rank all A_NCRC options by CVaR_α(L(a));

2. Select the option with minimal CVaR_α;

3. Require a time-bound mitigation plan with specific risk-reduction commitments and enhanced monitoring;

4. Document the TRC deficit in the PCC;

5. Trigger mandatory review based on severity classification;

6. Require one-tier-higher approval (Tier 3 decision requires Tier 4 oversight).

7.6 Catastrophic Pathological Unions and Emergency Ethics

Section 3 distinguished constructive from pathological unions and introduced the containment principle, while Sections 6 and 7 introduced the NCRC and the TRC as lexicographic safety layers. In most cases, misalignment can be handled within this ordinary regime: options that violate rights are discarded by NCRC, options with intolerable catastrophic tails are discarded by TRC, and the remaining option set is ordered by RLS and, when needed, the UCI. However, there are rare configurations where the union itself becomes a structural source of catastrophic risk. This subsection formalizes how MathGov treats such cases.

MathGov defines pathological unions as unions whose behavior systematically pushes ripple outcomes toward rights violations or TRC breach for containing unions, even when evaluated over rolling windows and under NCAR learning. Constructive unions may occasionally generate harm or risk, but their long-run pattern remains compatible with NCRC and TRC; pathological unions do not. Among pathological unions, catastrophic unions (CUs) are those for which continuation or scaling generates a non-negligible probability mass in the catastrophe corridor defined in Section 7, particularly at the CMIU and biosphere levels.

For emergency ethics, MathGov distinguishes two subtypes:

Non-adversarial Catastrophic Union (NCU). Catastrophic primarily through negligence, design error, or unsafe coupling, not through intrinsic goals. Examples include a poorly governed but well-intentioned geoengineering regime, or an AI-driven financial system that amplifies instability.

Adversarial Catastrophic Union (ACU). Catastrophic because its central objectives entail large-scale rights violations or extinction of other unions, such as a totalizing political regime that seeks permanent domination, or a digital system whose goal structure includes extinguishing or enslaving other intelligences.

NCUs and ACUs both trigger the catastrophe analysis in Section 7, but their ethical treatment differs. For an NCU, the default stance is remedial: the goal is to redesign, decouple, or de-scale the union so that it re-enters the TRC corridor while preserving as many of its legitimate interests and rights as possible. Emergency measures such as forced decommissioning or temporary suspension of certain operational freedoms are permitted only when no rights-respecting alternatives can bring the system back within bounds.

For an ACU, the rights calculus changes. A union whose explicit purpose is to extinguish or permanently subjugate other unions cannot be treated as morally equivalent to its intended victims. Within MathGov, NCRC is defined over all unions within the evaluative scope, and the rights floor of many unions collectively has lexicographic priority over the claimed interests of a single union that aims to destroy them. In such cases, it is ethically permissible to override the ACU's claim to continued operation or expansion, provided that emergency actions themselves remain subject to tightly constrained cruelty minimization and proportionality requirements.

At a procedural level, emergency ethics for catastrophic unions is still implemented via the existing cascade:

1. **Diagnosis.** Use NCAR and kernel-aware monitoring to detect whether a union's behavior matches the CU profile, including persistent TRC pressure and rights violations across multiple unions.

2. **Classification.** Distinguish NCU from ACU based on goal structure, revealed preferences, and evidential robustness, using independent audits where possible.

3. **Emergency option set.** Generate an option set restricted to measures that neutralize catastrophic risk while minimizing rights and welfare losses, prioritizing reversible and non-destructive interventions in NCU cases.

4. **Constrained selection.** Apply NCRC and TRC to the emergency option set itself, then rank admissible emergency options by RLS and UCI, with additional scrutiny on long-term union coherence.

5. Sunset and review. Treat emergency measures as temporary. Declare an explicit expiry and exit conditions in the PCC. If the emergency decision is made at Tier 1–2, require escalation to a Tier 3–4 retrospective review once the acute condition stabilizes, and no later than the next NCAR cycle, to confirm rights compliance, validate assumptions, and unwind measures that are no longer necessary.

This structure answers a key edge case: a fully extinction-seeking AI or regime is not treated as "misaligned but tolerated." Once classified as an ACU, it enters the emergency branch where its rights claims can be overridden to protect the rights and continued existence of other unions, while still binding responders to MathGov's non-compensatory rights logic in the design and execution of neutralization strategies. The formal criteria and decision rules for this branch are provided in Appendix P.

8. Ripple Propagation: The Kernel and Epistemic Humility

8.1 Why Ripples Must Be Modeled

Most real-world actions are not confined to a single union-dimension cell. A decision that improves Organization-Material might increase Community-Material via jobs and local spending, decrease Community-Environment via pollution, decrease Biosphere-Environment via emissions, and over time impact Humanity-Health through climate-related harms.

If we only model the direct impact in the initiating cells, we systematically mis-estimate consequences. To handle this, MathGov uses a ripple kernel **K** that encodes how changes in one cell propagate to others.

8.2 The Kernel Matrix

Let the 49 cells be indexed as pairs (u, d) with u ∈ {1,…,7} and d ∈ {1,…,7}. The kernel K is a sparse 49 × 49 matrix with entries:

interpreted as:

1. $\kappa > 0$: positive impact in source cell (u, d) tends to produce positive impact in target cell (u', d').

2. $\kappa < 0$: positive impact in source cell (u, d) tends to produce negative impact in target cell (u', d') (or vice versa).

3. $\kappa = 0$: no modeled systematic pathway between those cells.

Let i = ϕ(u, d) and j = ϕ(u', d') where ϕ(u, d) = 7(u − 1) + d. Kernel convention (canonical). K_{ij} maps effects from source cell j to target cell i (target-row, source-column). This convention ensures propagation uses standard left-multiplication by K in the equations below (e.g., Ĩ^prop = Ĩ^dir + K Ĩ^dir).

In practice, most entries are zero. The design target is sparsity. A memory cap can be enforced (e.g., a maximum number of non-zero entries per row, default ≤ 10, total non-zero entries ≤ 200) for interpretability and to avoid overfitting. Key Operational Pathways Set (KOPS) entries are exempt from trimming.

Kernel Semantic Interpretation. The kernel **K** represents a local linearization of ripple propagation around baseline conditions:

1. **Marginal effects: Each entry K_{ij} estimates the marginal effect of a unit change in the source cell j on the target cell i, holding other cells constant, evaluated at baseline conditions.**

2. **Sign behavior: Propagation is linear in the signed impact. If a positive unit change in source cell j produces K_{ij} change in target cell i, then a negative unit change in source cell j produces −K_{ij} change in target cell i.**

3. **Validity regime:** The linear approximation is valid within a bounded regime around baseline. The tanh saturation (Section 5.3 and 8.3.4) constrains both inputs and outputs to [−1, +1], ensuring the linear approximation operates within its validity domain.

4. **Scenario conditioning:** When stress conditions alter structural relationships (e.g., climate tipping points change Biosphere→Humanity pathways), scenario-conditioned kernels **K**_s should be used, documented in the PCC.

8.3 Propagation Equations (Kernel-Based Ripple Dynamics)

MathGov models indirect consequences ("ripples") by propagating direct impacts through a sparse ripple kernel **K**. Propagation is implemented in two stages: (i) linear propagation producing a pre-saturation propagated vector Ĩ^prop, and (ii) post-propagation saturation returning impacts to the normalized [−1, +1] scale.

8.3.1 Vectorization and Notation

Let U = {1,…,7} be the set of seven operational unions and D = {1,…,7} be the set of seven welfare dimensions. Let ϕ: U×D → {1,…,49} be a flattening map from cell (u, d) into a vector index.

Canonical flattening (default). Unless otherwise stated, MathGov uses row-major flattening by union then dimension:

where unions are ordered $U_1 < U_2 < \ldots < U_7$ and dimensions are ordered $D_1 < D_2 < \ldots < D_7$. Any deviation must be declared in the PCC.

For an option a, let the direct impact matrix be I^dir_{u,d}(a). Define the flattened direct impact vector:

whose components correspond to the 49 union-dimension cells.

Let K be the sparse ripple kernel. Under the canonical kernel convention, K_{ij} encodes how an impact in source cell j contributes to the propagated impact in target cell i.

8.3.2 Quick Mode (First-Order Propagation)

In Quick mode, MathGov uses a single propagation step:

This includes direct impacts plus first-order ripples. It is computationally efficient and is the default when kernel quality is low or stability checks fail.

8.3.3 Full kernel propagation (resummed mode) and stability requirements

Full kernel propagation models ripple effects beyond first-order spillovers by resumming indirect effects through the ripple kernel . This mode is intended for Tier 4 analysis when sufficient evidence exists to justify kernel structure and stability.

8.3.3.1 Inputs and notation

Let:

1. $I^{dir}(a) \in [-1,1]^{49}$be the flattened direct impact vector for option a, pre-saturation.
2. $K \in \mathbb{R}^{49\times49}$be the ripple kernel K, where K_{ij} maps effects from source cell j into target cell i (target-row, source-column).$K_{ij} j i$
3. I_{49}be the 49×49 identity matrix.

8.3.3.2 Full resummed propagation

Under the linear ripple approximation, total propagated (pre-saturation) effects satisfy:

$$\tilde{I}^{prop}(a) = I^{dir}(a) + K\,\tilde{I}^{prop}(a).$$

Rearrange:

$$(I_{49} - K)\tilde{I}^{prop}(a) = I^{dir}(a).$$

Thus:

$\tilde{I}^{prop}(a) = (I_{49} - K)^{\{-1\}} \cdot \tilde{I}^{dir}(a)$

Invertibility / stability note (Normative for Full mode). Full propagation is permitted only when $(I_{49} - K)$ is invertible and the run passes the declared stability gate (e.g., $\rho(K) < 1$ or a sufficient norm bound). If invertibility/stability cannot be established, the run MUST fall back to Quick mode or K = 0 per §8.3.3.6 and must record the fallback in the PCC.

Kernel convention reminder (audit-critical). K maps effects from source cell j to target cell i (target-row, source-column). Implementations MUST assert this convention in the PCC to prevent transpose bugs.

This expression captures direct effects plus all higher-order ripple paths K^2, K^3, ...implicitly.

8.3.3.3 Existence and convergence condition

Full mode requires that (I - K) is invertible. A sufficient stability condition is:

$$\mathrm{rho}(K) < 1$$

$$\rho(K) < 1,$$

where rho(K) is the spectral radius of K.

Operationally, MathGov uses a conservative sufficient bound:

$$||K||\_\mathrm{infty} < 1$$

$$\| K \|_{\infty} := \max_i \sum_j | K_{ij} | < 1,$$

$$(I - K)^{\wedge}(-1) = \mathrm{sum}\_\{n=0\}^{\wedge}\{\mathrm{infty}\}\ K^{\wedge}n$$

which implies convergence of the series and guarantees $(I - K)^{\wedge}(-1)$ exists.

$$\rho(K) \leq \| K \|_{\infty} < 1.$$

The PCC must report the stability check used and its computed value.

8.3.3.4 Post-propagation saturation

Because the resummed propagation can yield components outside $[-1,1]$, MathGov applies elementwise saturation:

$$\bar{I}^{prop}(a) := \tanh(\beta_{prop}\, \tilde{I}^{prop}(a)),$$

where $\beta_{prop} > 0$is the propagation saturation coefficient (default $\beta_{prop} = 1$). Then:

$$\bar{I}^{prop}(a) \in [-1,1]^{49}.$$

8.3.3.5 Documentation requirements for Full mode

Full mode is only valid if the PCC includes:

1. the kernel K(or a stable hash and retrieval link) and its provenance,
2. stability check results (ρ bound or norm bound),
3. the propagation mode used (Full vs Quick),
4. β_{prop}used,
5. any masking or sparsification rules applied to K,
6. justification that kernel entries are not double-counting direct impacts already represented in I^{dir}.

8.3.3.6 Fallback rules when stability fails

If stability fails ($\| K \|_{\infty} \geq 1$ or other diagnostics indicate instability), Full mode is not permitted. The methodology must fall back in this order:

1. **Quick mode**: $\tilde{I}^{prop}(a) = I^{dir}(a) + KI^{dir}(a)$
2. **No-kernel fallback**: set $K = 0$and proceed with $\tilde{I}^{prop}(a) = I^{dir}(a)$

Any fallback must be declared in the PCC, including why Full mode was rejected and what limitations the fallback introduces.

8.3.4 Post-Propagation Saturation (Required)

Because linear propagation can amplify values beyond [−1, +1], MathGov applies a second saturation step elementwise. Define Ĩ^prop_{u,d}(a) as the component of the pre-saturation propagated vector corresponding to cell (u, d). Then:

where β_prop is the post-propagation saturation coefficient (default β_prop = 1).

Equivalently, in vector form:

Canonical rule (tiered). NCRC checks use Ī^rights derived from Ī^prop (worst-off subgroup). TRC checks use L_raw(a,s) from AF-BASE/AF-EXT when Tier ≥ 4 (trc_mode = raw_indicator); bounded-impact TRC using Ī^prop(a|s) is permitted only for Tier ≤ 3 as declared, and is diagnostic-only for Tier ≥ 4. RLS ranking uses Ī^prop.

8.3.5 Scenario-Conditioned Propagation (for TRC and Scenario-Aware Analysis)

When evaluating tail risk under scenarios *S*, propagation is performed per scenario. Scenario dependence may enter via:

1. scenario-conditioned kernels **K**_s (if causal couplings change under stress), and/or
2. scenario-conditioned direct impacts **I**^dir(a | s) (if direct impacts differ by scenario).

The default is to keep **K** fixed and vary only the scenario conditions used to generate direct impacts. If **K**_s is scenario-conditioned, it must be justified and documented in the PCC.

For each scenario *s*, compute:

Quick mode:

*Full mode (requires ρ(**K**_s) < 1):*

and then:

Scenario-conditioned propagated impacts are used for subgroup rights evaluation (NCRC) and for scenario-aware RLS/diagnostics. Tier ≥ 4 admissibility TRC MUST use raw-indicator loss L_raw(a,s) from AF-BASE/AF-EXT; any bounded-impact TRC derived from Ī^prop is diagnostic-only and permitted only at Tiers ≤ 3 as a declared mode (see §7.2–7.3).

8.3.6 Interval Propagation (When Impacts Are Bounds)

When direct cell impacts are represented as intervals [I^lo, I^hi], MathGov requires an explicit interval-propagation rule.

Tier 3 default (auditable): (i) for NCRC/TRC, propagate the pessimistic endpoint for protected cells (the endpoint that maximizes violation depth or catastrophic loss) to avoid understating downside under uncertainty; (ii) for RLS, propagate midpoints (I^lo + I^hi)/2 and carry half-widths (I^hi − I^lo)/2 into the uncertainty treatment.

Tier 4 option: propagate both endpoints using sign-aware interval-matrix bounds.

The chosen interval rule must be recorded in the PCC. (If the kernel itself is interval-valued, see Section 8.5.)

8.4 Stability Constraints and Fallback Rules

MathGov imposes kernel stability constraints:

Entry bounds. Each non-zero entry satisfies |K_{ij}| ≤ κ_max (default κ_max = 0.5), preventing any single pathway from amplifying more than half the source signal in one step.

Absolute row-sum ($\boldsymbol{\ell}^{\mathbf{1}}$-norm) constraint. For all rows *i*, $\Sigma_j |K_{ij}| \leq \rho_{max}$ (default $\rho_{max} = 0.9$), which implies $\rho(\mathbf{K}) \leq \rho_{max} < 1$ and supports convergence in Full mode.

Spectral radius bound. Require $\rho(\mathbf{K}) < 1$ for Full mode. If $\rho(\mathbf{K}) \geq 1$, fall back to Quick mode or rescale **K** to enforce stability, and record the rescaling in the PCC.

Condition number check. Compute $\text{cond}(\mathbf{I} - \mathbf{K})$. If cond > 1000, the system falls back to Quick mode and logs this fallback in the PCC.

Non-zero entry cap. The number of non-zero entries is capped at ≤ 200 for interpretability and to avoid overfitting. KOPS entries are exempt from this cap.

K = 0 fallback behavior. If Full mode is infeasible (stability check fails) and Quick mode is deemed insufficient for the decision context, the system may fall back to K = 0 (no ripple propagation). This fallback must be: (i) explicitly recorded in the PCC as a "Kernel-Humility flag," (ii) accompanied by a statement that ripple effects are not modeled and the decision may underestimate cross-union consequences, and (iii) flagged for priority kernel calibration in the next NCAR cycle. K = 0 fallback is not permitted for Tier 4 decisions without governance escalation.

K = 0 Fallback Threshold:

When Kernel Quality Score falls below threshold:

KQS Policy (Single Source of Truth) (Normative).

Let KQS ∈ [0,1] be the Kernel Quality Score for the kernel used in the run. The following policy governs whether propagation is permitted and what constraints apply:

See KQS Policy (Single Source of Truth) in §8.5 for the authoritative thresholds and required actions.

- KQS ≥ 0.50: Kernel quality is acceptable for the declared tier, subject to other tier requirements.

All other mentions of KQS thresholds in this paper are subordinate to this policy.

1. KQS ∈ [0.40, 0.50): **K** = **0** fallback is recommended; if kernel is used, sensitivity analysis required
2. KQS ≥ 0.50: Kernel use permitted with standard sensitivity analysis

These constraints embody a Kernel Humility Principle: it is better to be explicitly conservative and incomplete in modeling ripples than to pretend to know more than we do.

8.5 Kernel Quality Score (KQS) and Interval Kernels

The propagation kernel K is an explicit model of cross-cell ripple effects. Its outputs MUST never be treated as more certain than the evidence supporting its edges.

MathGov assigns a Kernel Quality Score (KQS) in [0,1] to each kernel profile used in a run. KQS is an auditable summary of kernel readiness for decision-relevant propagation, and MUST be recorded in the PCC together with component scores, weights, the kernel convention (target-row, source-column), the propagation_mode used (None, Quick, Full), and any fallbacks triggered by this section.

KQS is computed from four 0-1 component scores: coverage C_cov (share of non-zero edges with evidence links), identifiability C_id (edges are replayable from specified endpoints, signs, and indicators), stability margin C_stab (conservative stability score), and predictive accuracy C_pred (out-of-sample pilot/backtest accuracy when available).

KQS := (w_cov·C_cov) + (w_id·C_id) + (w_stab·C_stab) + (w_pred·C_pred), where the component weights w_* are taken from the KQS Policy (Single Source of Truth) in this section.

Component	Weight
Coverage (w_cov)	0.25
Identifiability (w_id)	0.30
Stability (w_stab)	0.20
Prediction (w_pred)	0.25

Default component weights (provisional): w_cov = 0.25, w_id = 0.30, w_stab = 0.20, w_pred = 0.25. PCC records any deviations.

KQS Coverage Hardening (Tier 4). If C_cov < 0.80 for the relied-upon edge set (defined below), cap KQS <= 0.49 and set audit_flag KERNEL_COVERAGE_INCOMPLETE.

This section is the single source of truth for KQS bands, tier overrides, and required actions. Any other KQS guidance in the paper is subordinate unless it is identical to this section.

8.5.1 KQS Policy (Single Source of Truth) (Normative)

(A) KQS bands and required actions

KQS < 0.40: Kernel MUST NOT be used for decision-relevant propagation. propagation_mode MUST be None (Ī^prop := I^dir). The kernel may be retained as a research artifact but MUST NOT affect NCRC, TRC, containment, or RLS. Starter-KOPS profiles shipped in ProofPack SHOULD include a declared KQS in the kernel profile registry; if KQS is absent, implementations MUST treat KQS as 0.00 and set propagation_mode = NONE for Tier ≥ 4.

0.40 <= KQS < 0.50: Kernel use is sensitivity-gated. Only Quick propagation is permitted (never Full). The PCC MUST run kernel sensitivity (±0.05 on relied-upon edges or the declared interval half-width, whichever is larger). If admissibility changes or the selected option flips under permitted perturbations, escalation is REQUIRED.

0.50 <= KQS < 0.65: Kernel use is permitted. Quick is permitted at Tier 3 and Tier 4. Full is permitted only at Tier 4 and only if stability gates pass. The PCC MUST run kernel sensitivity on relied-upon edges and attach an evidence bundle for relied-upon edges. Set audit_flag KQS_MEDIUM.

KQS >= 0.65: Kernel use is encouraged. Quick is permitted. Full is permitted only at Tier 4 and only if stability gates pass. Evidence bundling and tier constraints still apply.

(B) Tier overrides and hard constraints (always apply)

Tier 3 MUST NOT use Full propagation, regardless of KQS. Tier 4 (Pilot-Executable, rev14.x) MUST NOT use Full propagation. Implementations MUST hard-fail a Tier-4 Pilot-Executable claim that sets propagation_mode=FULL (audit_flag FULL_PROPAGATION_PROHIBITED_TIER4_REV14). If a non-Tier-4 run requests FULL but the implementation does not support FULL, it MAY fall back to QUICK or NONE and MUST record the fallback in the PCC.

If KQS is capped by KERNEL_COVERAGE_INCOMPLETE, the run is restricted to the 0.40-0.50 band rules above, even if the uncapped KQS would be higher.

(C) No silent K = 0

If the kernel is disabled due to KQS band rules, coverage hardening, or failed stability gates, the PCC MUST explicitly record propagation_mode = None and MUST include a brief limitation statement (kernel humility note) describing what ripple pathways are not being modeled.

Tier defaults (rev14.1): Tier-1 and Tier-2 default propagation_mode = None. Tier-3 defaults to propagation_mode = Quick (KOPS) unless the PCC explicitly sets propagation_mode = None. Tier-4 defaults to propagation_mode = Quick (KOPS) unless the PCC explicitly sets propagation_mode = None. Any override from the tier default MUST be declared in the PCC and treated as decision-relevant configuration.

(D) Relied-upon edge definition (for sensitivity requirements)

An edge K_ij != 0 is relied-upon if perturbing it by ±0.05 (or by its interval half-width, if interval-valued) can change any of: NCRC admissibility, TRC pass/fail, containment pass/fail, or the top-ranked option within the declared judgment threshold δ. Implementers may conservatively treat all non-zero edges as relied-upon. The PCC MUST record the relied-upon edge set and the perturbation rule used.

8.5.2 Computing KQS component scores (Tier 3 starter rubric)

This subsection specifies a minimal, repeatable method for computing component scores. Alternative rubrics are permitted, but MUST compute each component on a 0-1 scale and MUST be documented in the PCC.

Let E be the set of non-zero kernel entries (i,j) (edges), with |E| its size. If |E| = 0, set C_cov = C_id = C_stab = 1, and set C_pred = 0.50 unless outcome-tracking evidence exists.

(1) Coverage C_cov: C_cov = #{(i,j) in E : evidence(i,j) != empty} / |E|, where evidence(i,j) is the set of evidence items linked to edge (i,j) in the PCC evidence log.

(2) Identifiability C_id: C_id = #{(i,j) in E : identifiable(i,j) = true} / |E|, where identifiable(i,j) means the PCC records source cell, target cell, sign, and at least one measurable indicator family for each endpoint cell.

(3) Stability C_stab: let ||K||_infty = max_i Σ_j |K_ij|. Define C_stab = clamp_[0,1](1 - max(0, ||K||_infty - 0.90) / 0.10). Tier 4 Full mode additionally requires the stability gates in §8.4.

(4) Predictive C_pred: let m be the number of evaluated decisions with recorded kernel-affected outcomes. If m < 10, set C_pred = 0.50 and mark it as a low-evidence prior. If m >= 10, compute C_pred = clamp_[0,1](1 - MAE / MAE_ref). The PCC MUST declare the evaluation set, outcome definition, and scoring method.

(5) Evidence-class floor (Tier 4): if more than 50% of non-zero edges are Weak evidence class (Class E), cap C_pred <= 0.60 unless pilot/backtest evidence justifies lifting the cap under governance.

8.5.3 Interval kernels (Normative)

When point estimates are uncertain, MathGov allows interval-valued kernels: K_ij ∈ [K_ij^lo, K_ij^hi]. For NCRC and TRC, the run MUST use pessimistic evaluation, applying the kernel setting within the declared intervals that yields the worst admissibility outcome per option, and recording the bound rule used. For RLS ranking, the PCC MUST report an RLS interval per option induced by the kernel intervals. If rankings overlap or flip within the interval range, the run MUST be treated as sensitivity-dominated and handled as a judgment call under the declared δ threshold, with escalation if required by tier policy.

8.6 Key Operational Pathways Set

Not all kernel entries are equally important. For transparency and governance, MathGov defines a Key Operational Pathways Set (KOPS): the subset of 50-150 entries that have documented empirical or theoretical justification, account for a large share of RLS variance, and represent load-bearing causal pathways.

Sources for KOPS entries include Health Impact Assessment literature, Integrated Assessment Models for climate (Nordhaus, 2017), input-output economic tables (Leontief, 1986), network science on social propagation (Christakis & Fowler, 2009), ecological food web models (Estes et al., 2011), and the Marmot Review on social determinants of health (Marmot, 2010).

Sensitivity analysis tools, including one-at-a-time perturbations and Sobol indices, identify which kernel entries have the highest marginal effect on RLS, prioritize data collection and model refinement, and highlight where disagreements or uncertainty have the greatest ethical significance.

A Starter KOPS with literature-derived entries is provided in Appendix S for organizations beginning implementation.

8.7 Kernel Validation Gate

Before claiming "pilot-ready" status for any kernel profile, the following validation must be completed:

Retrospective Sign-Accuracy Test:

1. Assemble a test set of at least 100 historical decisions with known outcomes
2. For each KOPS entry, predict the sign of propagated impact
3. Compare predictions to observed outcomes
4. Compute sign accuracy:

Minimum Threshold: Sign accuracy ≥ 0.60 for the kernel to be approved for Tier 4 use

Calibration Error Test:

For entries with sufficient data, compute mean absolute error between predicted and observed impact magnitudes:

Maximum Threshold: MAE ≤ 0.25 for the entry to remain in KOPS without "high uncertainty" flag

Pre-Validation Disclaimer:

Until Phase 2 validation completes, all decisions using Starter KOPS must include PCC disclaimer:

"Ripple predictions based on unvalidated literature estimates. Starter KOPS coefficients have not undergone retrospective validation against observed outcomes. Uncertainty bounds are expanded by 50%. This disclaimer will be removed upon completion of Phase 2 validation with sign accuracy ≥ 0.60."

9. Sentience and Rights: The Sentience Gradient Protocol (SGP)

9.0 Normative Authority and Integration Rule (SGP 4.1.1)

Sentience determination for all rights-impacting decisions in MathGov 5.0i rev14.29 SHALL use the canonical Sentience Gradient Protocol (SGP) v4.1.1 as the exclusive normative protocol. Any earlier or alternative sentience heuristics that appear in this Foundation Paper are non-authoritative and SHALL NOT be used to assign rights floors, tier classifications, or sentience multipliers for Tier-4 Pilot-Executable computation.

MathGov consumes SGP outputs only through the official binding defined in the MathGov Appendices, Appendix G (SGP 4.1.1). The canonical scalar mapping is:

- **SGP_score(E) := min(A(E), B(E), C(E))**, where each pillar score is in [0,100]

- **SG_norm(E) := SGP_score(E) / 100**, where SG_norm(E) ∈ [0,1]
- **s_k := SG_norm(E)** is the sentience multiplier used by MathGov when a sentience multiplier is required

Normative plateau rule for human persons:
For rights protection and safety, all human persons are treated as full rights-plateau stakeholders. Therefore, for any human person H, set **SG_norm(H) := 1.0** by normative commitment, independent of measurement noise, uncertainty, or partial observability.

When SGP classification is relevant to a MathGov decision instance, evaluators SHALL follow the steps and artifact outputs required by SGP v4.1.1 and Appendix G.

9.1 What SGP Governs Within MathGov

SGP governs moral patienthood classification and related rights-of-protection activation. It is the normative mechanism for determining whether an entity is a credible welfare-bearing stakeholder and which minimum protections apply under the Non-Compensatory Rights Constraint (NCRC).

SGP governs the following elements of the MathGov Operating System:

- Classification of moral patienthood tiers (SGP-0 through SGP-5)
- Confidence bands, evidence class reporting, and measurement conservatism requirements
- Activation of minimum rights-of-protection floors under the NCRC
- Stability and robustness requirements for higher-tier claims (including evaluation windows and adversarial resilience)
- Required record fields for Tier-4 auditability when sentience affects permissible actions

This Foundation Paper provides only an integration summary. For any decision, audit, or publication claim that depends on sentience classification, SGP v4.1.1 is authoritative and supersedes any earlier in-document sentience criteria, thresholds, or tests that conflict with the SGP protocol.

9.2 Rights Are Not Authority

SGP determines rights-of-protection and constraints on treatment. It does not automatically grant governance power, ownership privileges, or decision authority. Authority eligibility is separately gated within MathGov by competence, alignment, auditability, non-domination, and revocability requirements. This separation prevents governance capture while still protecting any entity with credible welfare-bearing status.

9.3 Required Decision Record Fields When Sentience Is Relevant

Whenever sentience classification affects a decision, the decision record SHALL include, at minimum:

- Entity identifier(s) and scope of interaction
- SGP tier assignment (SGP-0 to SGP-5)
- Confidence band and evidence class
- SG_norm(E) value and the rule used (including the human plateau rule if applicable)
- Rights-of-protection floor activation and any resulting constraints under NCRC
- Evaluator identity, method summary, and artifacts required by Appendix G

In closed-access systems where internal telemetry, architecture, or reliable behavioral exposure is unavailable, SGP classifications SHALL be conservative. Evaluators may require explicit ceilings, wider uncertainty bands, or precautionary handling without asserting consciousness.

9.4 Integration Hooks (Appendix G Binding)

Appendix G (SGP–MathGov Integration Note) is the authoritative binding layer for translating SGP outcomes into permitted actions and governance constraints. When SGP is invoked, the evaluator SHALL:

- Apply NCRC rights floors consistent with the SGP tier (rights floors cannot be traded away for aggregate welfare gains)
- Apply any constraint escalations required by Appendix G
- Record the tier, confidence, evidence class, and SG_norm(E) in the decision record
- Preserve the required artifacts for audit replay where Tier-4 execution is claimed

9.5 Legacy Diagnostic Heuristic (Non-Normative, Deprecated for Canon Use)

The following component-weight heuristic is retained only as a diagnostic aid for discussion and sensitivity checking. It is not a canonical method and SHALL NOT be used to compute SG_norm(E), assign SGP tiers, infer rights floors, or determine Tier-4 permissible actions. Canonical sentience determination is exclusively defined by SGP v4.1.1 and Appendix G.

Component	Weight	Description	Evidence Class
Neural/Computational Complexity	0.15	Structural capacity for information integration	Moderate
Behavioral Indicators	0.25	Observable responses suggesting experience	Strong
Self-Referential Processing	0.20	Stable internal state representation and meta-cognition	Moderate
Affective Responses	0.20	Ability to experience pleasure, pain, stress, satisfaction	Strong
Meta-Cognitive Indicators	0.10	Awareness of own mental states	Weak
Integrated Information	0.10	Phi or equivalent measures of integration	Weak

Note: This table is informational only. It SHALL NOT be used as a substitute for SGP v4.1.1 evaluation or Appendix G binding rules.

10. Weights and Value Aggregation: Hybrid Democratic Weighting

10.1 The Problem of Weights

The RLS requires weights over unions (w_u) and dimensions (v_d). These weights encode value judgments about relative importance. Who decides them?

Two failure modes must be avoided. Pure technocracy, where experts impose weights, undermines democratic legitimacy and invites accusations of elite bias. Pure democracy, where majorities can vote to zero out protections, enables majority tyranny over minorities, future generations, and the environment.

Hybrid Democratic Weighting (HDW) is designed to navigate between these extremes by combining constitutional floors with democratic tuning.

10.2 Floors as Constitutional Constraints

MathGov defines floors for union and dimension weights that cannot be violated in any implementation. These floors are set based on structural necessity (e.g., Biosphere as planetary substrate), locus of experience (Self), and interdependence across unions.

Union weight floors:

Union	Floor (w^floor_u)	Rationale
Self	0.20	Locus of experience and agency
Household	0.06	Primary care/resource unit
Community	0.06	Local social fabric
Organization	0.06	Productive coordination
Polity	0.08	Public goods and governance
Humanity/CMIU	0.10	Species-level coordination
Biosphere	0.10	Planetary life-support
Total	**0.66**	Constitutional minimum

Dimension weight floors:

Dimension	Floor (v^floor_d)	Rationale
Material	0.08	Basic needs
Health	0.10	Biological viability
Social	0.08	Relational integrity
Knowledge	0.08	Epistemic capacity

Dimension	Floor (v^floor_d)	Rationale
Agency	0.10	Self-determination
Meaning	0.06	Existential orientation
Environment	0.10	Ecological sustainability
Total	**0.60**	Constitutional minimum

These floors ensure that no union or dimension can be mathematically eliminated from consideration. The Self floor is highest because individual agents are the locus of experience. Biosphere and CMIU floors are elevated because they represent long-term substrate conditions.

10.3 Structural and Democratic Components (Hybrid Democratic Weighting)

HDW decomposes each weight into (i) a structural component grounded in evidence about interdependence, systemic risk, and cross-union effects, and (ii) a democratic component grounded in deliberative preference formation. HDW is applied separately to union weights w_u and dimension weights v_d.

Let w^floor be the union floor vector and v^floor be the dimension floor vector, with Σ_u w^floor_u = 0.66 and Σ_d v^floor_d = 0.60. Floors are constitutional constraints that prevent any union or welfare dimension from being zeroed out through preference aggregation, lobbying, or capture.

Let w^str and v^str be structural proposal vectors on the simplex, and w^dem, v^dem be democratic proposal vectors on the simplex derived from deliberative democratic processes.

10.3.1 Tier 4 structural proposal vectors (w^str, v^str): internal default and derivation

Purpose. Tier 4 requires a usable, auditable structural baseline without relying on packaged charters. This subsection defines a minimal, internal procedure and a canonical default for the structural proposal vectors w^str (unions) and v^str (dimensions). Implementations may replace these with a charter-set procedure at Tier 4, but any change must be declared in the PCC with a rationale and sensitivity check.

Tier 4 canonical default (use if no structural data are available). Use the following simplex vectors:

Union structural proposal (w^str, Tier 3 starter):

Self 0.10; Household 0.10; Community 0.12; Organization 0.12; Polity 0.14; Humanity/CMIU 0.20; Biosphere 0.22.

Dimension structural proposal (v^str, Tier 3 starter):

Material 0.20; Health 0.20; Social 0.12; Knowledge 0.12; Agency 0.12; Meaning 0.12; Environment 0.12.

Minimal derivation procedure (auditable; optional for Tier 4). If structural evidence is available, derive w^str and v^str as follows:

(1) Start with the Tier 3 starters above (or uniform vectors on the simplex).

(2) Apply multiplicative multipliers based on declared, auditable indicators, then renormalize: w^str_u ∝ w^0_u · (1 + κ_U · Z_u) and v^str_d ∝ v^0_d · (1 + κ_D · Z_d), with κ_U=0.25 and κ_D=0.25 by default.

(3) For unions, Z_u is the mean of up to three normalized indicators in [0,1]: (a) exposure scale (stakeholder count or affected population share), (b) externality reach (how far impacts propagate beyond the union), and (c) irreversibility horizon (typical time horizon of impacts in that union).

(4) For dimensions, Z_d is the mean of up to two normalized indicators in [0,1]: (a) rights adjacency (fraction of canonical rights whose coverage sets include any cell in dimension d, per Appendix C.3.7) and (b) measurement reliability (inverse of typical uncertainty for that dimension in the given context).

(5) Floors and simplex constraints. w^str and v^str must satisfy the simplex constraints and MUST NOT violate floor feasibility. If a derived value falls below a floor, clamp to the floor and renormalize the remaining mass across non-clamped elements; record the clamping event in the PCC.

Output. Record in PCC: whether defaults or derived values were used, indicator definitions and sources, κ values, any clamping events, and the final w^str and v^str vectors.

Define blend parameters λ_U and λ_D, where λ is the democratic share of the above-floor mass (the remainder follows the structural proposal). Default values are λ_U = 0.70 and λ_D = 0.70 unless otherwise set by governance, meaning that 70% of above-floor mass follows democratic allocation and 30% follows the structural proposal.

HDW blend semantics (Single Source of Truth) (Normative).

Let w_floor be the constitutional floor vector over unions with floor_mass := Σ_u w_u^floor and allocable_mass := 1 − floor_mass. Let w_dem and w_str be proposal vectors on the union simplex (Σ_u w_u^dem = Σ_u w_u^str = 1, w_u^dem ≥ 0, w_u^str ≥ 0). Let λ_U ∈ [0,1] be the democratic share of the allocable mass. Then the final union weights are:

w_u := w_u^floor + allocable_mass · (λ_U · w_u^dem + (1 − λ_U) · w_u^str).

Analogously for dimensions: with v_floor, allocable_mass_D := 1 − Σ_d v_d^floor, v_dem, v_str, and λ_D as the democratic share of allocable dimension mass:

v_d := v_d^floor + allocable_mass_D · (λ_D · v_d^dem + (1 − λ_D) · v_d^str).

By construction, Σ_u w_u = 1 and Σ_d v_d = 1, and floors are always satisfied. Any deviation from these formulas MUST be treated as a nonstandard HDW variant and MUST be explicitly declared and justified in the PCC.

10.4 Anti-Capture and Integrity Mechanisms in HDW

Because weights steer optimization, they are a primary target for capture. MathGov therefore treats weight-setting as a high-integrity governance function and requires safeguards that are auditable, adversarially tested, and revision-controlled.

Stratified deliberation and representation. HDW assemblies must include stratified representation across unions and stakeholders, including at minimum: household/community representatives, organizational stakeholders, polity-level governance representatives, biosphere advocates or scientific stewards, independent risk experts, and vulnerable population representatives. Stratification rules (selection, rotation, and conflict-of-interest checks) are specified in the PCC.

Minimum representation requirements. Each union type must have at least one delegate in the HDW assembly. Vulnerable population representatives (including representatives for future generations where feasible) must be included.

Biosphere Steward Requirement (Normative). Any HDW assembly for decisions with material Environment (D_7) impacts at Polity/Humanity/Biosphere scales MUST include at least one Biosphere Steward delegate with a formal mandate to advocate for U_7. The PCC MUST include the steward's mandate basis and a conflict-of-interest disclosure. Material financial ties to directly benefiting parties require recusal.

Supermajority locks near floors. Any proposal that reduces a union or dimension weight to below a guarded proximity band (default: floor + 0.02) requires a supermajority threshold (default: 2/3) in the HDW assembly, plus an independent review panel sign-off. This prevents incremental erosion of protected weights through repeated small changes.

Transparency ledger and immutability. All weight proposals, vote tallies, dissenting statements, and rationales are published in a public ledger with immutable hashes. The published record must include the floor vector, structural and democratic proposals, blend parameters, and the resulting computed weights, enabling third-party verification.

Algorithmic red-teaming and back-testing. Before adoption, proposed weights are tested against a reference suite of historical and synthetic decisions. The test suite is designed to detect systematic distortions, such as chronic underweighting of biosphere outcomes, repeated rights-near misses, or increased tail-risk exposure. Test results and identified failure modes are logged in the PCC.

Protected-minority and cultural-harm trigger. If a culturally distinct subgroup can demonstrate that a weight revision would predictably cause rights violations or irreparable cultural harm within protected dimensions, the revision triggers a pause-and-mediation protocol. The protocol must include evidence review, facilitated negotiation, and a formal written resolution. The goal is not to grant arbitrary veto power, but to prevent erasure and unaccounted harms that standard aggregation can miss.

Conflict-of-interest and funding disclosure. All participants in weight-setting and review panels must disclose material conflicts (financial, institutional, political). When conflicts exceed PCC thresholds, recusal is mandatory. Panel composition and recusals are logged.

These mechanisms are not optional; they define the minimum integrity posture for HDW and are treated as part of system feasibility, not merely best practice.

10.5 Floor Governance Charter and Amendment Procedure

Floors are constitutional, not immutable. They may be revised, but only through a high-friction amendment procedure designed to prevent capture, prevent rights erosion, and preserve cross-context comparability.

A floor revision proposal must include, at minimum:

1. The exact floor changes proposed (current and proposed values, by union and dimension).
2. A justification grounded in empirical evidence and system objectives.
3. A risk assessment explicitly addressing rights exposure, tail-risk effects, and potential gaming incentives.
4. A transition plan, including monitoring signals and rollback triggers.

Adoption requires dual authorization:

1. A supermajority vote in the representative HDW assembly (default: 2/3), and
2. A supermajority vote in an independent review panel (default: 2/3), whose mandate is to assess capture risk, rights integrity, and tail-risk exposure.

Review cadence. Floors are reviewed on a fixed cadence (default: every 36 months) and may be reviewed earlier only under strong new evidence or major systemic change. Cadence and triggers are defined in the PCC.

Global minima and local tightening. Global floors define minimum protections. Local contexts may tighten floors (e.g., increasing biosphere protection in a stressed ecosystem), but may not loosen floors below global minima without passing the full amendment procedure above. All local deviations must preserve non-compensatory rights constraints.

Full transparency. All proposals, rationales, evidence packets, votes, dissenting opinions, and final decisions are published and hashed. No floor revision is valid unless its PCC entry is complete and auditable.

10.6 Cultural Localization and Measurement Invariance

MathGov measures outcomes (needs fulfillment and welfare dimensions), not cultural strategies. Different cultures may pursue different pathways to reach similar welfare outcomes, and the framework must be compatible with pluralism while retaining global comparability for rights protections.

Localization is permitted and encouraged in indicators, semantics, and measurement instruments, subject to four requirements.

Local indicator proposals. A locale may propose local indicators and semantic interpretations for each dimension, provided the mapping to the canonical dimension definition is explicit and documented in the PCC.

Measurement invariance testing. Localization must include evidence that the proposed indicators measure the intended construct in a comparable way across relevant populations. At minimum, locales must test for basic invariance (e.g., stability of interpretation under translation, response consistency, and construct validity checks) and publish the results.

Orthogonality preservation. Dimensions are designed to be approximately separable for decision analysis. Localization must verify that dimensions remain approximately non-redundant, using a governance-defined criterion (default: |r| < 0.85 between dimension measures over the local reference set), or else document why redundancy is acceptable and how it is managed.

Rights-floor comparability. For rights-covered cells, floors must retain stable meaning across contexts. Localization may change indicators, but it may not reinterpret a rights floor as "less severe" by rescaling. For rights-constrained cells, the anchoring reference class and mapping must be declared so that a floor violation corresponds to the same category of real-world harm across locales.

All localization mappings, calibrations, invariance tests, and audit artifacts are recorded in the PCC.

11. Scoring and Selection Under Uncertainty

11.1 Ripple Logic Score

The Ripple Logic Score (RLS) aggregates weighted impacts across all cells:

$$\mathrm{RLS}(a) := \sum_{u \in U} \sum_{d \in D} w_u \ v_d \, \bar{I}_{u,d}(a), \qquad \bar{I}_{u,d}(a) := \sum_{s \in S} p_s \ I_{u,d}(a,s)$$

where w_u and v_d are HDW-weighted union and dimension weights, m_{u,d} is the applicability mask, and E_S[Ī^prop_{u,d}(a)] is the expected post-propagation, post-saturation impact over scenario set S:

Sentience scaling is applied within-cell when aggregating entity-level impacts into I^dir (Sections 9.4-9.5), not as an additional union-level multiplier.

Scenario-set default. When scenario-aware RLS is used, *S* defaults to the same governed scenario set used for TRC evaluation. A distinct scenario set for RLS may be used only with explicit justification and PCC documentation (for example, shorter-horizon welfare scenarios versus long-horizon catastrophe scenarios).

11.2 Uncertainty Propagation

Because impacts, kernel entries, and weights may carry uncertainty, MathGov tracks an approximate RLS uncertainty σ_RLS. A conservative first approximation treating cell-level uncertainties as independent:

$$\sigma_{\mathrm{RLS}}(a) := \sqrt{\sum_{u \in U} \sum_{d \in D} \left(w_u \, v_d \, \sigma_{u,d}(a)\right)^2}$$

where σ_{u,d}(a) is the recorded cell-level impact uncertainty (e.g., half-width of the interval or a calibrated standard deviation proxy).

More sophisticated implementations can incorporate covariance between cells and kernel uncertainty.

This uncertainty estimate is epistemic rather than frequentist: it reflects the decision-maker's confidence bounds on impact estimates given available evidence, not a sampling distribution from repeated trials. The independence assumption is intentionally conservative; implementations with richer covariance information can represent impacts, kernel entries, and weights as a joint uncertainty structure and propagate it through the kernel using standard multivariate methods, yielding a more faithful picture of the overall state of knowledge.

11.2.1 Default cell-level uncertainty mapping (Tier 4)

Tier 4 requires a computable default method to derive sigma_{u,d}(a) from the instance records in K(u,d,a). Unless the PCC declares an alternative uncertainty model, use the following rules.

Rule A (interval-first). If a cell impact is recorded as an interval [I_lo_{u,d}(a), I_hi_{u,d}(a)], set:

sigma_{u,d}(a) = (I_hi_{u,d}(a) - I_lo_{u,d}(a)) / 2

This is a conservative proxy (half-width). The PCC must state if a different conversion is used.

Rule B (confidence-based when no interval is provided). If only impact instances with confidence c_k in [0.1,1] are provided, define an instance uncertainty proxy:

sigma_k = clip((1 - c_k) * |mu_k| , 0 , 1)

and define normalized instance weights:

omega_k = (r_k * tau(t_k) * ell_k) / (sum_{j in K(u,d,a)} r_j * tau(t_j) * ell_j)

Then compute the cell uncertainty as:

sigma_{u,d}(a) = sqrt(sum_{k in K(u,d,a)} omega_k^2 * sigma_k^2)

If K(u,d,a) is empty, set sigma_{u,d}(a) = 0 only when the cell is explicitly declared measured-zero with supporting evidence recorded in the PCC. For Tier 4, an active cell (m_{u,d}=1) MUST NOT be empty unless it is measured-zero; otherwise the PCC is INVALID (ACTIVE_CELL_EMPTY_INSTANCE_SET_INVALID).

Kernel uncertainty note (Tier 4). If kernel uncertainty is not modeled, treat K as fixed and compute sigma_RLS from cell uncertainties only. If K is interval-valued, the PCC must state whether it uses endpoint propagation or a bounded perturbation method, and must report the resulting sigma_RLS.

This default mapping makes the discrimination-band logic operational without requiring packaged statistical modeling, while remaining conservative and auditable.

11.3 Risk-Adjusted RLS

When decision-makers prefer to penalize options with high uncertainty more strongly, MathGov offers a risk-adjusted RLS:

$$\mathrm{RLS}_{\mathrm{adj}}(a) := \mathrm{RLS}(a) - \lambda\, \sigma_{\mathrm{RLS}}(a)$$

with λ ≥ 0 (default λ = 0.5) controlling the degree of risk-aversion.

11.4 Discrimination Threshold and Judgment Calls

11.4 Discrimination threshold, uncertainty, and “judgment call” triggers

MathGov ranks admissible options using the Ripple Logic Score $RLS(a)$. In practice, however, differences in RLScan be too small relative to uncertainty, measurement noise, and modeling error to justify a confident selection. This section defines a **single canonical discrimination test** and a governance-safe trigger for judgment calls and tie-break escalation.

11.4.1 Discrimination band via the gap function

For any two admissible options $a, b \in A_{adm}$, define the normalized separation:

$$gap(a,b) := \frac{|\ RLS(a) - RLS(b)\ |}{\sigma_{RLS}(a) + \sigma_{RLS}(b) + \epsilon},$$

where:

1. $\sigma_{RLS}(a) \geq 0$is the estimated uncertainty (standard deviation) of $RLS(a)$under the selected uncertainty model (Section 11.2), and
2. $\epsilon > 0$is a small stabilizer constant to prevent division instability (default $\epsilon = 0.01$).

Interpretation. $gap(a,b)$measures how large the difference in scores is relative to the combined uncertainty. Large gap implies a robust separation; small gap implies the apparent difference is within noise.

11.4.2 Canonical discrimination rule

Let a^*be the highest-scoring admissible option under RLS. Let b^*be the second-highest.

Define a governed discrimination threshold $\Delta_{disc} > 0$(default $\Delta_{disc} = 1.0$). Then:

1. **Decisive lead:** If $gap(a^*, b^*) \geq \Delta_{disc}$, then the RLS ordering is treated as decisively separable for selection purposes (subject to containment in §11.6).

2. **Non-decisive lead (judgment band):** If $gap(a^*, b^*) < \Delta_{disc}$, then the top two options are within the discrimination band, and the selection must proceed using the tie-break and escalation rules in §11.5–§11.6.

The PCC must report $gap(a^*, b^*)$, the chosen Δ_{disc}, and the uncertainty model used to compute σ_{RLS}.

11.4.3 When $\boldsymbol{\sigma_{RLS}}$is unavailable

Default weights (if not specified). Tier 1–2: if the PCC does not specify w_U or w_D, use uniform defaults. Tier 3: uniform defaults are permitted only as an explicitly declared fallback when HDW ballots are unavailable, and the PCC MUST set an audit_flag indicating HDW fallback and record the rationale. Tier 4: uniform defaults are NOT permitted; w_U and w_D MUST be derived from a valid HDW ballots registry, otherwise the tier claim MUST be downgraded. The PCC MUST set audit_flag = TIER_CLAIM_DOWNGRADE_REQUIRED.

$$gap(a,b) = \frac{\mid RLS(a) - RLS(b) \mid}{\epsilon}.$$

In this case, the discrimination threshold Δ_{disc}must be interpreted as a pure minimum-difference rule. The PCC must explicitly state that uncertainty was not modeled and that the discrimination rule is therefore less robust.

11.5 UCI, HOI, and structural coherence tie-breaks

When RLS differences are non-decisive, MathGov uses structural coherence metrics to prevent selecting an option that appears welfare-superior but weakens the integrity, resilience, or fairness of the underlying unions. This section defines the Union Coherence Index (UCI), the Hollowing-Out Index (HOI), and the canonical tie-break chain.

11.5.1 Union Coherence Index (UCI): definition and components

Indicator operationalization boundary (Normative).

Appendix E provides canonical indicator families and examples for UCI components (cohesion, flow, resilience, equity). These are intended as reference families, not as universally validated measurement instruments.

Validated operational protocols (e.g., specific survey items, scoring rubrics, reliability/validity coefficients, and measurement invariance results) are deployment- and context-specific and may be developed progressively through NCAR (Reflect) and the validation program in §14.

Implementations MUST document their indicator operationalization choices (and any evidence of reliability, construct validity, and invariance where applicable) in the PCC, and MUST treat unvalidated instruments as PROVISIONAL for audit and certification purposes.

Tier ≥ 4 UCI input contract (Normative). For replayability, the PCC MUST include the normalized component inputs x_H[u], x_F[u], x_R[u], x_E[u] for each union u as exact rationals, and MUST either (i) hash-bind a derivation protocol artifact (data source + normalization transform) or (ii) explicitly declare that the component inputs were panel-provided judgments under a named procedure. UCI_V1 MUST be computed solely from these declared component inputs.

For each union $u \in U$, define component scores:

$$H_u, F_u, R_u, E_u \in [0,1],$$

representing:

1. H_u: cohesion (internal connectivity / trust / alignment),
2. F_u: flow (functional throughput and coordination),
3. R_u: resilience (shock tolerance and recovery capacity),
4. E_u: equity (fair distribution of burdens/benefits and voice).

Let component weights $\gamma_H, \gamma_F, \gamma_R, \gamma_E \geq 0$satisfy:

$$\gamma_H + \gamma_F + \gamma_R + \gamma_E = 1.$$

Then define:

$$UCI_u := \gamma_H H_u + \gamma_F F_u + \gamma_R R_u + \gamma_E E_u.$$

Define the aggregate UCI across unions using governed union weights $v_u \geq 0$with $\sum_u v_u = 1$:

$$UCI := \sum_{u \in U} v_u \ UCI_u.$$

Default weights (if not specified). If the PCC does not specify $v$or γ, use uniform defaults.

Tier 3 structural-independence requirement. At Tier 4, the indicators used to compute must be structural/process indicators distinct from the welfare matrix used for RLS. Appendix E defines canonical indicator families by union and component.

11.5.2 Prospective (ex ante) UCI estimation from structural indicators (Tier 3)

For Tier 3 decisions, UCI must support prospective evaluation of options through forecasted structural changes, not welfare-to-UCI proxy mapping.

Let baseline component levels $H_u^{base}, F_u^{base}, R_u^{base}, E_u^{base} \in [0,1]$be recorded. For each option a, estimate bounded component changes:

$$\Delta H_u(a), \Delta F_u(a), \Delta R_u(a), \Delta E_u(a) \in [-1,1]$$

using the same impact-instance pipeline as §5.2–§5.4 (indicator anchoring, aggregation, saturation), but with the structural indicators listed in Appendix E.

Compute:

$$\Delta UCI_u(a) := \gamma_H \Delta H_u(a) + \gamma_F \Delta F_u(a) + \gamma_R \Delta R_u(a) + \gamma_E \Delta E_u(a).$$

Update levels with clipping:

$$UCI_u(a) := \operatorname{clip}(UCI_u^{base} + \Delta UCI_u(a), 0, 1),$$

and then compute aggregate $UCI(a)$via the union weights v_u.

Documentation requirement. Tier 3 PCCs must list the structural indicators used for each component and their anchoring reference classes.

$\Delta UCI_u(a) \bar{I}_{u,d}^{prop}(a)$Prohibited shortcut at Tier 3. Deriving directly from welfare impacts is prohibited at Tier 3 because it collapses UCI into a re-aggregation of RLS.

Tier-3 UCI Unavailability Rule (Required). If structural indicators per Appendix E are unavailable such that UCI cannot be computed without violating Tier-3 structural independence, then UCI MUST be treated as unavailable for tie-break purposes. In this case:

(i) if the top candidates are within the RLS discrimination band, the decision MUST escalate to additional data collection and/or a higher tier, or
(ii) a documented governance "judgment call" may be made only with explicit PCC labeling JUDGMENT_CALL_UCI_UNAVAILABLE, including rationale and monitoring plan.

Any welfare-derived UCI proxy MUST NOT be used to claim Tier-3 compliance.

11.5.3 Hollowing-Out Index (HOI): definition on differences

HOI is a diagnostic used in ongoing monitoring to detect a pattern where welfare scores improve while coherence erodes. HOI is defined on **changes**, not raw levels.

Let $RLS_i$and UCI_ibe values at review period i. Define first differences:

$$\Delta RLS_i := RLS_i - RLS_{i-1}, \Delta UCI_i := UCI_i - UCI_{i-1}.$$

Let $Smooth(\cdot)$be the exponential moving average (EMA) with smoothing parameter $\lambda \in (0,1]$:

$$Smooth(x)_i := \lambda x_i + (1-\lambda) Smooth(x)_{i-1}.$$

Then define:

$$HOI_i := Smooth(\Delta RLS)_i - Smooth(\Delta UCI)_i.$$

Interpretation.

1. Persistent $HOI > 0$indicates welfare scores improving faster than coherence, suggesting hollowing risk.
2. Persistent $HOI < 0$indicates coherence improving faster than welfare scores.

HOI is not an admissibility filter. It is an audit and monitoring diagnostic used in PCC follow-up, governance review cycles, and containment escalation when appropriate.

11.5.4 Canonical tie-break chain when RLS is non-decisive

When $gap(a^*, b^*) < \Delta_{disc}$, the following tie-break chain applies, in order:

1. **Containment priority (pre-selection gate).** Evaluate containment (Mode A) in RLS order per §11.6. If the top candidate fails containment and is rejected/escalated, evaluate the next candidate.
2. **UCI dominance.** Prefer the option with higher $UCI(a)$if the difference exceeds a governed UCI discrimination threshold Δ_{UCI}(default $\Delta_{UCI} = 0.05$).
3. **HOI risk flag (if monitoring context exists).** If one option is associated with persistent positive HOI under comparable monitoring conditions, treat it as riskier and escalate or prefer the alternative, depending on governance policy.
4. **Escalation trigger.** If no clear winner emerges, invoke a judgment-call protocol (documented in the PCC) or escalate to a higher tier / additional analysis.

The PCC must record which step resolved the tie or why escalation occurred.

11.6 Containment check, selection procedure, and escalation rules

Containment prevents selecting an option that passes NCRC and TRC yet threatens structural integrity or creates brittle, gameable, or destabilizing conditions. Containment is treated as an **integrity gate prior to final selection**.

11.6.1 Containment concept and modes

Containment can be implemented in different modes depending on tier and available information, but the governance meaning is constant: a selection must not introduce unacceptable structural failure conditions.

1. **Mode A (Canonical containment gate).** A deterministic pass/fail (or pass/escalate) rule based on governed containment indicators and thresholds. Mode A is the canonical integrity gate used in the cascade.
2. **Mode B (Monitoring containment).** A longitudinal containment audit over time, using UCI/HOI diagnostics and other structural indicators to detect drift and gaming. Mode B supports post-decision governance, not the immediate selection step.

Unless otherwise declared, Tier 4 uses Mode A for selection.

11.6.2 Canonical containment selection algorithm (RLS-ordered evaluation)

Let A_{adm} be the admissible set after NCRC and TRC. Let $\prec$ denote ordering by decreasing RLS. Define the ordered list:

$$a_{(1)} \succ a_{(2)} \succ \cdots \succ a_{(k)},$$

where $a_{(1)}$ has the highest RLS.

Selection proceeds as follows:

Guard (Normative): Selection is Mode A–only. The selection run MUST be executed under Containment Mode A. Mode B outputs are diagnostic-only and MUST NOT be used as inputs to selection, ranking, tie-break, or escalation decisions. If a PCC shows Mode B affected selection, the PCC MUST be labeled INVALID with audit_flag CONTAINMENT_MODE_B_USED_FOR_SELECTION.

1. Compute $RLS(a)$ for all $a \in A_{adm}$.
2. Order candidates by RLS from best to worst.
3. For $j = 1$ to k (in RLS order):
   3.1) Evaluate **Containment (Mode A)** for candidate $a_{(j)}$.
   3.2) If containment **passes**, mark $a_{(j)}$ as the current selection candidate and proceed to the discrimination and tie-break rules (Step 4).

3.3) If containment **fails**, do not select $a_{(j)}$. Record the failure mechanism in the PCC and either:

1. reject $a_{(j)}$and continue to $a_{(j+1)}$, or

2. escalate immediately if the failure indicates structural hazard requiring governance review (Step 6).

4. If the top two containment-pass candidates are decisively separated ($gap \geq \Delta_{disc}$), select the highest RLS containment-pass option.

5. If not decisively separated ($gap < \Delta_{disc}$), apply §11.5 tie-break chain among containment-pass candidates.

6. If no containment-pass option exists (or containment failures indicate systemic hazard), escalate per §11.6.4.

This procedure makes containment an integrity gate while preserving computational efficiency by evaluating containment only in RLS order until a pass is found or escalation occurs.

11.6.3 What constitutes a containment failure (required declaration)

Containment failure criteria must be defined in the PCC for Tier 4 and are mandatory at Tier 4. At minimum, containment must declare thresholds for at least one of the following classes:

1. **Structural fragility:** increased single points of failure, reduced redundancy, elevated systemic coupling beyond safe ranges.

2. **Governance capture risk:** evidence the option increases manipulability of weights, masking, indicators, or kernel entries.

3. **Coherence collapse risk:** projected significant decline in UCI components beyond a governed floor or beyond an acceptable negative change bound.

4. **Path dependence and lock-in:** creation of irreversible dependency, centralization, or incentive structures likely to entrench misalignment.

Appendix E provides default indicator families for coherence-related containment criteria. If additional containment indicators are used, they must be declared, anchored, and bounded.

11.6.4 Escalation rules

Escalation is triggered when:

1. $A_{adm} = \emptyset$after NCRC and TRC (handled earlier by emergency/fallback protocols), or

2. no containment-pass candidate exists, or

3. containment failures reveal a hazard that requires higher-tier governance review, or

4. RLS is non-decisive and UCI tie-breaks are also non-decisive.

When escalation occurs, the PCC must specify one of the following actions:

1. Raise tier (Tier 4 → Tier 3, Tier 3 → Tier 4) and rerun with stronger uncertainty and structural indicators.

2. **Expand options** (generate new alternatives) and rerun the cascade.

3. **Modify constraints** only if governance permits, documenting the rationale and the tradeoffs explicitly.

4. **Emergency ethics protocol** if urgency overrides standard deliberation (must be explicitly labeled and logged).

Escalation must never silently bypass NCRC or TRC. Any exceptional decision mode must be declared as such.

11.7 Containment Governance Parameters

The Containment Principle (Section 3.4) and its operationalization (Section 11.6) introduce several governance parameters that must be specified in the PCC:

Parameter	Symbol	Default Value	Allowed Range	Governance Level
Containment tolerance	τ_c	−0.10	[−0.20, 0.00]	Charter (global); PCC (tightening only)
Positive-impact threshold	θ_pos	0.05	[0.01, 0.10]	PCC
Containment depth limit	D_c	2	{1, 2, 3}	PCC
Containment mode	Mode A/B	Mode A	—	PCC (Mode B requires explicit justification and cannot enable selection)

Parameter governance rules:

τ_c (containment tolerance): The global default of −0.10 may be tightened (made less negative or positive) for critical containing unions such as Biosphere. It may not be loosened below −0.10 without Charter-level revision. A tolerance of −0.10 means that a containing union's coherence may decline by up to 10% before triggering containment failure; tightening to −0.05 would require coherence to decline by no more than 5%.

θ_pos (positive-impact threshold): Determines which unions are considered "positively impacted" and therefore subject to containment checks. Lower values are more conservative (more unions checked). Values below 0.01 are discouraged as they may trigger containment checks for noise-level positive impacts, creating computational overhead without meaningful ethical benefit.

D_c (containment depth): Limits the ancestral chain checked for coherence degradation. Default of 2 checks immediate container and its container. Higher values increase computational cost and may be appropriate for Tier 4 decisions with complex nesting structures. For most decisions, D_c = 2 captures the most ethically significant containing unions without excessive computation.

Containment mode: Mode A (veto/escalation) is required for all Tier 4 decisions. Mode B (disqualification) is permitted only for Tier 2 exploratory analysis with explicit governance approval and cannot be used to enable selection of containment-violating options. A PCC that invokes Mode B for a selected option is automatically flagged for audit.

Ancestor mapping Anc(u, D_c). The set of containing unions for union *u* up to depth D_c is determined by the standard nesting hierarchy:

For any union *u*, Anc(u, D_c) returns the next D_c unions in this chain. For example:

1. $Anc(U_1, 2) = \{U_2, U_3\}$

11.7A Mode A Containment Gate Algorithm (Normative; single source of truth)

This block is the authoritative operational definition of Containment Mode A. If any other passage conflicts, this block governs.

Inputs (from PCC snapshots + registries):

• Candidate option a; propagation_mode ∈ {NONE, QUICK}; and per-cell post-propagation impacts Ī^{prop}_{u,d}(a) already computed under NDP_FIXEDPOINT_V1.

• UCI_u(a) and ΔUCI_u(a) computed per Appendix E.7, using the declared UCI component set and fixed-point checkpoints.

• Containment parameters: D_c, θ_pos, τ_c (from hash-bound registry or PCC override).

Step 1 — Identify positively-moving unions (U_pos):

Compute S_u(a) = Σ_d v_d · Ī^{prop}_{u,d}(a) using the declared dimension weights v_d and post-propagation impacts. Define U_pos(a) = {u | S_u(a) ≥ θ_pos}.

Step 2 — Compute Mode A containment minima:

For each u ∈ U_pos(a): let A_u = Anc(u, D_c) (ancestor set per §11.7). Compute M_u(a) = min_{v ∈ A_u} ΔUCI_v(a).

Step 3 — Gate condition (PASS/FAIL):

Containment_ModeA_Pass(a) = TRUE iff for all u ∈ U_pos(a), M_u(a) ≥ τ_c.

Step 4 — Normative action in Tier-4 selection:

If Containment_ModeA_Pass(a)=FALSE, option a MUST be removed from the selectable set prior to final selection.
Normative set definitions: A_adm := {a ∈ A | NCRC(a)=PASS ∧ TRC(a)=PASS}. A_sel := {a ∈ A_adm | Containment_ModeA_Pass(a)=TRUE}.
Selection rule: the chosen option MUST be argmax_{a∈A_sel} RLS(a) (with the published tie-breaks). Implementations MAY compute RLS only for A_sel for efficiency, but MUST be able to reproduce the RLS values that would have been obtained for A_adm.
If A_sel is empty, the run MUST declare CONTAINMENT_EMPTY_SELECTABLE_SET and either revise the option set or revise containment parameters under a new PCC revision (or accept a tier downgrade for that run).

Required transcript checkpoints (Tier-4 replay):

Record, in this exact order: U_pos(a); each A_u; each M_u(a); τ_c; θ_pos; and the final boolean Containment_ModeA_Pass(a).

2. Anc(U_4, 2) = {U_5, U_6}

3. Anc(U_6, 2) = {U_7}

4. Anc(U_7, 2) = ∅

When a union has fewer than D_c ancestors in the hierarchy, Anc(u, D_c) is the set of all available ancestors. If Anc(u, D_c) is ∅, then that u contributes no containment minima (i.e., M_u(a) is vacuously PASS for that u).

12. The NCAR Learning Loop

12.1 Motivation

Even a formally precise decision system can fail if inputs are stale, biased, or mis-specified. NCAR is the required learning loop that updates assumptions, weights, and registries based on measured outcomes and observed failure modes.

12.2 The Four Stages

Notice. Define the decision scope and option set *O*. Map affected unions and dimensions. Set the applicability mask m_{u,d} with documented rationale, verifying non-maskable cells are included. Identify relevant rights thresholds, tail-risk scenarios, and load appropriate kernel profile. Establish baselines for key indicators and UCI. The Notice stage culminates in a structured specification suitable for both analysis and future audit.

Choose. Run the lexicographic cascade following the Canonical Impact Construction Algorithm (Section 3.2.7). Estimate direct impacts and propagate ripples via **K**. Apply NCRC with worst-off subgroup checks, TRC, and containment check. Compute RLS and uncertainty; identify Judgment Calls when intervals overlap; apply UCI/HOI tie-breaks. Enforce Containment Principle. Select an option and record predicted impacts, expected UCI trajectory, and key assumptions with confidence levels.

Act. Implement with monitoring aligned to assumptions used in Choose. Track indicators used to estimate impacts in real time or at agreed intervals. Monitor TRC scenarios for early warning signals. Track structural metrics for UCI. Record implementation variance, distinguishing between errors in the model and errors in execution.

Reflect. Compare observed outcomes to predictions. For each key cell and indicator, compute hit rates (proportion of impacts where observed sign or magnitude fell within predicted bands). Check whether any rights were violated in practice even if NCRC predicted admissibility. Check whether realized tail losses exceeded TRC expectations. If systematic prediction errors are found, adjust magnitude calibrations, update kernel entries (especially in KOPS), modify coefficients for UCI. Propose revisions to floors, thresholds, or HDW settings if evidence suggests mis-calibration, routing such proposals through the Floor Governance Charter. Document all lessons, adjustments, and rationales in a versioned registry, cross-referenced to the PCC of the original decision.

12.3 NCAR Across Tiers

NCAR applies at all implementation tiers. What changes by tier is the minimum evidence and audit depth required in Reflect.

Tier 1 (Heuristic). Reflect is informal and qualitative (for example, journaling, brief notes on harms avoided, and what was learned).
Tier 2 (Core, Calculable). Reflect uses simple before-after checks on a small set of declared indicators and verifies that no rights-floor violations were missed.
Tier 3 (Standard). Reflect uses structured data collection, basic trend checks, and compares observed outcomes against the declared assumptions (including any Starter-KOPS predictions if used), recording any recalibration decisions.
Tier 4 (High Assurance). Reflect uses rigorous evaluation (model comparison when applicable, sensitivity analysis, and where feasible formal tests of predictive claims), and updates registries/kernels only through the declared governance pathway and version control.

Default timing for Reflect phase:

1. Tier 3 decisions: Reflect within 6 months of implementation or after significant outcome data becomes available.

2. Tier 4 decisions: Reflect within 3 months or after significant outcome data becomes available.

3. Emergency Mode decisions: Reflect at review intervals specified by severity classification (Section 6.4).

In all cases, the principle remains: MathGov is not a one-shot oracle; it is a continuously improving system.

12.4 Kernel Temporal Validity and Structural Change Detection

The kernel **K** is a model of ripple propagation under current structural conditions. When structural relationships change significantly, **K** may become invalid. MathGov includes protocols for detecting and responding to kernel invalidity:

Structural change indicators. The following signals may indicate that **K** requires re-evaluation:

1. Major regime changes (political, economic, technological) affecting modeled pathways
2. Observed outcomes systematically diverging from kernel-based predictions (hit rate < 0.50 over 10+ decisions)
3. Identified tipping points or phase transitions in relevant systems
4. Expert assessment that key causal relationships have changed

Kernel validity review trigger. A kernel validity review is triggered when:

1. Cumulative prediction error for KOPS-pathway outcomes exceeds 30% across a rolling window of decisions
2. A major structural change event occurs (documented and classified in PCC)
3. Regular review cadence is reached (default: 24 months)

Kernel update procedure. When a validity review is triggered:

1. Identify which kernel entries are affected by the structural change
2. Assess whether new evidence supports revised coefficients
3. If updates are warranted, update entries with documented justification
4. Version the kernel (K-v1.0 → K-v1.1) and record change log
5. Re-run sensitivity analysis for pending decisions under old and new kernels
6. Archive old kernel; do not delete (for audit purposes)

Interim measures. While kernel validity is under review:

1. Use conservative (reduced) kernel coefficients for affected pathways

2. Flag decisions as "kernel-validity-pending" in PCC

3. Consider Quick mode fallback for affected pathways

12.4.1 Kernel learning procedure (provisional, calculable default) (Normative).

Purpose. When NCAR (Reflect) identifies systematic kernel mis-specification (persistent prediction error patterns attributable to specific pathways), update kernel coefficients using an explicitly governed procedure so “update K” is calculable and reproducible.

Scope. This default procedure updates only KOPS edges (the declared operational pathway set). Off-pathway coefficients remain zero unless governance activates new edges via the kernel registry.

Data. For each completed decision run r, record predicted propagated impacts $\bar{I}^{pred}_r$ and observed realized impacts $\bar{I}^{obs}_r$ using the same cell mapping $\phi(u,d)$, scenario structure (if any), and baseline semantics.

Loss. Define a prediction loss L over a window of runs R as $L := (1/|R|) \cdot \Sigma_{r \in R} || \bar{I}^{pred}_r - \bar{I}^{obs}_r ||_2^2$ (or an alternative governed loss declared in the PCC).

Update rule (gradient step). For each KOPS edge (i,j), update $K_{ij}^{new} := \mathrm{clip}(K_{ij}^{old} - \gamma \cdot \partial L/\partial K_{ij} , -1, +1)$, with learning rate $\gamma \in (0,1]$. Default $\gamma := 0.10$.

Minimum data gate. Kernel updates MUST NOT occur until at least $N_{min} := 10$ completed decisions exist for the relevant domain class (or a governed alternative).

Governance and versioning. Any update MUST be emitted as a new hash-bound kernel registry version (e.g., REG-KERNEL-v1.0 → REG-KERNEL-v1.1) with a changelog (which edges changed, by how much, why), and MUST be referenced in the PCC for subsequent runs.

Audit note. This learning rule is PROVISIONAL and intended to prevent implementer invention. More advanced causal or Bayesian updating methods MAY be substituted only if declared in the PCC and governed as a registry-defined variant.

12.5 NCAR Calibration Metrics and Triggers

Sign Accuracy:

Trigger: If Sign Accuracy < 0.60 over 20+ decisions, initiate kernel recalibration.

Magnitude RMSE:

Trigger: If RMSE > 0.30 for any union-dimension cell across 20+ decisions, flag that cell for calibration review.

Rights Near-Miss Rate:

Trigger: If Near-Miss Rate > 0.25, review rights threshold calibration.

12.6 Inter-PCC Consistency

Multiple PCCs from the same organization should maintain consistency unless parameters have been explicitly updated. MathGov requires:

Configuration management. Organizations maintain a versioned configuration file specifying: kernel profile (version and entries), HDW weights, rights thresholds, TRC parameters, and any local governance adaptations. Each PCC references a specific configuration version.

Drift detection. If two PCCs from the same organization within a 12-month period use materially different parameters without an intervening governance update, this is flagged as "configuration drift" and triggers review.

Update propagation. When parameters are updated through governance procedures, a changelog is maintained and all subsequent PCCs reference the new version.

13. Provenance, Compliance, and Auditability

Artifact Integrity Law (AIL) and ProofPack Reliance (Normative)

13.1 Purpose (Normative)

MathGov is a shared operating system for decisions across unions. To preserve Union Coherence and prevent governance gaming, every nontrivial decision MUST be reproducible, comparable across time, and independently auditable. This requires tamper-evident registry referencing and a consistent Provenance and Compliance Certificate (PCC).

13.2 Definitions (Normative)

Registry. A versioned, governed set of normative or operational objects used by MathGov runs, including but not limited to rights coverage, rights anchors, catastrophe indicators, kernels, weights, and scenarios.

Registry Manifest. A machine-readable canonical representation of a registry whose content hash uniquely identifies its full contents and version.

PCC (Provenance and Compliance Certificate). The decision record artifact binding a run to (i) registry hashes, (ii) inputs, (iii) cascade outputs, and (iv) signoffs.

Content Hash. A cryptographic digest (default SHA-256) computed over canonicalized registry manifests or PCC content.

Embedded Snapshot. A minimal set of registry-derived values included inside the PCC to allow offline audit and human review even if registry retrieval is unavailable.

13.3 Core Registry Integrity Rules (Normative)

AIL1 (Registry Binding). Every PCC for Tier ≥ 2 MUST include:

- A hash reference to every registry artifact used by the run, and
- An embedded snapshot sufficient to independently recompute NCRC, TRC, and RLS for that decision.

No decision may claim MathGov compliance at any tier unless it is bound to explicit registry hashes.

AIL2 (Immutability). A registry referenced by hash in a PCC MUST be treated as immutable. Updates MUST create a new registry version with a new hash. Old versions MUST NOT be overwritten or deleted.

AIL3 (Comparability). Two options compared in the same decision MUST be evaluated under the exact same registry hashes (same rights coverage, anchors, catastrophe indicators, kernel, weights, scenario library) and the same tier and propagation mode configuration. If comparability cannot be met, the PCC MUST declare COMPARISON_INVALID, or MUST decompose into comparable subdecisions.

AIL4 (No Silent Overrides). Any override of a registry object MUST be expressed as either:

- A new registry version with a new hash (preferred), or
- A PCC-declared override bundle with its own hash, explicitly linked as an override layer on top of a base registry.

Silent ad hoc modification of thresholds, coverage sets, kernel entries, weights, or scenarios is prohibited.

13.4 Embedded Snapshot Requirements (Normative)

AIL5 (Offline Audit Sufficiency). Each PCC MUST embed the following snapshot fields at minimum:

For NCRC (rights admissibility):

- Effective rights thresholds θ_r used.
- Effective coverage sets C_r used (or a hash plus a fully expanded list).
- Subgroup policy and subgroup lists used for rights-covered cells.
- Worst-off subgroup impacts used for checks ($\bar{I}_{rights}$ values).

For TRC (tail-risk admissibility):

- Catastrophe cell set C_{cat} used.
- Catastrophe weights ω used.
- TRC parameters α and τ_{TRC}.
- Scenario IDs and probabilities p_s.
- Per-option loss values $L(a,s)$ and CVaR results.
- TRC mode (raw_indicator or bounded_impact).

For Ripple Propagation (if used):

- Propagation mode (None, Quick, Full).
- Kernel convention identifier (canonical target-row and source-column mapping).
- Kernel edges applied (edge IDs plus coefficients after scaling) or a kernel matrix hash.

For RLS:

- Union weights w_u and dimension weights v_d.
- Applicability mask m_{u,d}.

Snapshot values MUST match referenced registry manifests. Any mismatch triggers AIL7 (REGISTRY_MISMATCH).

Audit rule. If any narrative default conflicts with referenced registries, the PCC MUST include audit_flag DOC_DEFAULT_CONFLICT. The run remains valid only if it is replayable from the registries and PCC snapshots.

Normative Spine (Quick Map) (Informative)

To implement or audit this specification, treat the following as the minimal authoritative chain (in order):

5. Tier Requirements Matrix (§4.4.5) for tier gating and allowed modes.
6. HDW ballots and fixed-point weight derivation (HDW registries and rules in the Foundation and Appendix AA).
7. Impact computation pipeline (direct plus propagation), including KQS screening and allowed propagation modes.
8. Admissibility cascade: NCRC (rights) then TRC (tail-risk CVaR) then Containment Mode A.
9. Ranking over selectable set via RLS, with tie-break policy and audit flags.
10. Tier-4 determinism: NDP_FIXEDPOINT_V1 plus SAT_LUT_FP_V1 plus REG_TEMPORAL_WEIGHTS_V1 plus NO_FLOATS canonical JSON, all hash-bound in ProofPack.

When a PCC references hash-bound registries, registry contents are the single source of truth for all numeric and structural objects used in the run. Any narrative defaults, examples, or tables elsewhere in the Foundation or Appendices are non-binding guidance unless they are identical to the referenced registries.

13.5 Registry Precedence Rule (RPR) (Normative)

RPR1 (Registries Override Narrative). For any Tier ≥ 2 run, when a PCC references hash-bound registries, the referenced registries override any conflicting defaults, examples, or narrative parameter descriptions in the Foundation Paper or Appendices. Narrative content remains informative unless explicitly declared normative and proven identical to the referenced registries.

RPR2 (Single Source of Numeric Truth). All numeric parameters required to compute NCRC, TRC, Containment, Ripple Propagation, and RLS MUST be traceable to either (i) hash-bound registries referenced in the PCC, or (ii) a PCC override bundle with its own hash (AIL4). If a number cannot be traced, the run MUST NOT claim Tier-4 compliance and MUST be flagged NO_SOURCE_OF_TRUTH.

RPR3 (Governed Updates Only). Any change that would alter admissibility outcomes or rankings (for example, rights thresholds, catastrophe weights, scenario probabilities, weights, kernel edges, or indicator anchors) MUST be made via a governed new registry version. The PCC MUST record the previous version hash and the new version hash when such a change occurs.

13.6 PCC Validity and Audit Flags (Normative)

AIL6 (PCC Validity Predicate). A PCC is valid if and only if: (i) all required registry hash references are present, (ii) embedded snapshot values are consistent with referenced registries within declared numeric tolerances, and (iii) the PCC includes the required cascade trace (NCRC → TRC → Containment → RLS → tie-break and selection). If any condition fails, the PCC MUST be labeled INVALID and cannot be used to claim MathGov compliance.

AIL7 (Registry Mismatch Flag). If any embedded snapshot item differs from referenced registry content, the PCC MUST be flagged audit_flag REGISTRY_MISMATCH and MUST include: mismatched objects, expected hash or value, observed hash or value, and the responsible module identity.

AIL8 (Configuration Drift Flag). If materially different registry hashes are used across PCCs within an organization or deployment without a recorded governance update event, the system MUST flag audit_flag CONFIG_DRIFT and trigger review.

13.7 Tier 4 Pilot-Executable Determinism (Normative)

AIL9 (No Missing Numbers). For Tier 4 (Pilot-Executable) runs, every numeric parameter required to execute the run (thresholds, weights, scenario probabilities, kernel coefficients, raw-indicator anchors and mappings) MUST be present in hash-bound registries referenced in the PCC, or embedded in a PCC override bundle with its own hash (AIL4). No invention is permitted.

AIL10 (Mode Lock). For Tier 4 (Pilot-Executable) runs, TRC MUST be executed in raw_indicator mode using AFBASE catastrophe indicators unless the PCC declares an approved extension (C_ext plus AFEXT). Any bounded_impact TRC computation may be logged for diagnostics but MUST NOT determine admissibility.

AIL11 (Replay Test Duty). A Tier 4 (Pilot-Executable) PCC MUST pass a replay test: independent re-execution using PCC inputs and referenced registries reproduces admissibility results, ranking, and selected option within declared numeric tolerances. Failed replay invalidates the Tier 4 claim.

13.8 Numeric Determinism Profile (NDP) (Normative)

Purpose. Tier 4 Pilot-Executable claims require bit-stable numeric replay across independent implementations. This section defines the Numeric Determinism Profile (NDP) used for all Tier 4 computations that are not themselves hash-canonical artifacts.

13.8.1 Scope Split (Normative)

(A) Hash-canonical artifacts (registries, manifests, schemas, HDW ballots, HDW weights, PCC override bundles) MUST use exact reduced rationals of the form {num, den} with den > 0, and MUST conform to the canonical JSON profile and NO_FLOATS rule.

(B) Computed quantities (impact aggregation totals, saturation outputs, propagation outputs, losses, CVaR, RLS, discrimination gaps) MUST be computed under NDP_FIXEDPOINT_V1 and MUST be stored in the PCC as fixed-point integers only.

Tier-4 clarification. This includes UCI and ΔUCI values used for containment. Tier-4 PCC artifacts MUST record UCI_BASELINE_FP, UCI_OPTION_FP, and DELTA_UCI_FP, or schema-equivalent fields, as fixed-point int64 values under NDP_FIXEDPOINT_V1.

13.8.2 NDP_FIXEDPOINT_V1 (Normative)

Fixed-point scale. Let S = 10^9. Any real value x is represented as X_fp = round_half_even(x × S) as a signed int64. All divisions MUST use round-half-even. Tooling MUST hard-fail on overflow with audit_flag NUMERIC_OVERFLOW.

Rational-to-fixed conversion (canonical). For any exact rational {num, den}, convert by X_fp = round_half_even(num × S / den). This is the only permitted bridge from hash-canonical rationals to computed fixed-point values.

Saturation (Tier 4). Tier 4 MUST NOT compute tanh(·) or any floating approximation at runtime. Saturation MUST use a hash-bound fixed-point lookup table (LUT) in ProofPack: SAT_LUT_ID = SAT_LUT_FP_V1. The PCC MUST reference the LUT by hash and record SAT_LUT_ID.

Temporal weighting (Tier 4). Tier 4 MUST NOT compute logarithms at runtime. Tier 4 temporal weights MUST be sourced from a hash-bound registry in ProofPack: TEMPORAL_WEIGHT_REGISTRY_ID = REG_TEMPORAL_WEIGHTS_V1, then converted to fixed-point using the canonical conversion above.

Propagation mode restriction (Tier 4, rev14.x). For Tier 4 Pilot-Executable in rev14.x, propagation_mode MUST be NONE or QUICK. Full propagation is prohibited until a deterministic solver profile is released and hash-bound. If FULL is requested, tooling MUST hard-fail Tier 4 with audit_flag FULL_PROPAGATION_PROHIBITED_TIER4_REV14.

Quick propagation (fixed-point). When QUICK is used, compute I_prop_pre_fp = I_dir_fp + matmul_fp(K_fp, I_dir_fp) in fixed-point. Kernel entries MUST be stored as exact rationals in registries and converted entrywise to fixed-point via the canonical conversion.

Quantization checkpoints (required). Tier 4 tooling MUST quantize and store in the PCC: I_dir_pre_fp, I_dir_fp, I_prop_pre_fp, I_prop_fp, L_raw_fp[a,s], CVaR_fp[a], and RLS_fp[a].

Comparisons. All admissibility checks and tie-break comparisons MUST use exact fixed-point integer comparison with no epsilon, unless a tolerance is explicitly provided by a hash-bound registry and recorded in the PCC.

13.8.3 Tier 4 PCC Numeric Profile Fields (Normative)

Every Tier 4 PCC MUST include: numeric_profile_id = NDP_FIXEDPOINT_V1, S = 1000000000, SAT_LUT_ID with hash reference, TEMPORAL_WEIGHT_REGISTRY_ID with hash reference, propagation_mode, and if QUICK is used: kernel registry hash, kernel convention ID, and flattening map ID.

13.9 Governance Alignment and Redaction (Normative)

AIL12 (Union Coherence Through Shared Registries). Deployments SHOULD converge on shared registries at the appropriate union scale (organization, polity, CMIU), while allowing localized extensions through governed versioning. Local divergence MUST be explicit (hash-visible), justified, and reviewable.

AIL13 (Public Transparency vs Protected Data). PCC MUST support a Public PCC (redacted) and a Full PCC (protected). Redaction MUST NOT break integrity: both forms MUST carry hashes enabling auditors to verify faithful redaction. Redaction method MUST be declared in the PCC.

13.10 AIL Clause Index (Reference, Non-Normative)

AIL1 Registry Binding; AIL2 Immutability; AIL3 Comparability; AIL4 No Silent Overrides; AIL5 Offline Audit Sufficiency; AIL6 PCC Validity Predicate; AIL7 Registry Mismatch Flag; AIL8 Configuration Drift Flag; AIL9 No Missing Numbers; AIL10 Mode Lock; AIL11 Replay Test Duty; AIL12 Shared Registries; AIL13 Redaction Integrity.

13.11 The Provenance and Compliance Certificate (PCC) (Normative)

The Provenance and Compliance Certificate (PCC) is the primary artifact for accountability. It is a structured, machine-readable record that captures:

- Header: decision identifier, timestamp, decision owner(s), and spec version referenced.
- Scope: option set with descriptions, unions and dimensions in scope, m_{u,d} values and rationale for any exclusions, and non-maskable cell verification.
- Inputs: impact instances with μ, r, t, ℓ, c, and e values and sources; kernel profile identifier and KQS; rights thresholds and coverage sets applied; TRC parameters (α, τ_TRC, scenarios); weight profiles (w_u, v_d with HDW source).
- Cascade results: NCRC violations for each option with subgroup analysis; TRC CVaR values and pass or fail for each option; containment check results; RLS and σ_RLS for each option; final selection and rationale including tie-break logic and Judgment Call flags.
- Sensitivity analysis results (Tier 4): key parameter perturbation results; kernel sensitivity; weight sensitivity; threshold sensitivity; robustness classification (Robust, Sensitive, Fragile).
- Five-Sentence Public Rationale (5SPR): a five-sentence public rationale summarizing the decision in plain language, avoiding jargon.
- Signatures and hashes: content hash for integrity verification; sign-offs from responsible parties.

Option set edge cases (Normative).

- If the initial option set O is empty, the run MUST terminate as INVALID_INPUT_NO_OPTIONS and the decision owner MUST generate at least one feasible option before proceeding.
- If $|O| = 1$, the run MAY proceed for documentation, but MUST label SINGLE_OPTION_NO_SELECTION in the PCC and MUST treat the result as a constraint check (NCRC, TRC, Containment) rather than a comparative selection.

This single document enables internal accountability (who decided what, based on which inputs) and packaged audit (can independent reviewers reconstruct the reasoning).

13.12 Immutable Audit Ledger (Informative)

For high-stakes decisions such as large public policies, critical infrastructure, and AI system deployment, PCCs should be stored in an append-only ledger using Merkle-tree-based or equivalent structures, anchored in a tamper-evident system with cryptographic proofs that PCCs have not been altered post hoc.

This does not require all details to be public. Sensitive personal data may be redacted, but the structural integrity of the PCC must remain verifiable.

13.13 Role Separation (Informative)

MathGov emphasizes separation of roles to reduce conflicts of interest:

- Analysts: estimate impacts, define instances, calibrate kernels, and propose parameter settings. Must document methods, assumptions, and data sources.
- Decision owners: accountable for choosing among admissible options. Cannot silently alter analyst inputs; any overrides must be explicit in the PCC.
- Auditors: conduct independent reviews of PCCs. Check consistency of methods across decisions and investigate anomalies (for example, repeated borderline NCRC passes).

This structure mirrors three-lines-of-defense models in risk management and helps prevent MathGov from becoming mere ethics theater.

13.14 PCC Red-Teaming Protocol (Informative)

To proactively identify weaknesses, MathGov incorporates a formal red-teaming process for high-stakes PCCs:

- Independent Audit Lottery: 5% of PCCs are randomly selected for independent audit by a third-party ethics review board.
- Adversarial Scenario Testing: auditors construct worst-case scenarios not originally considered and re-run TRC and RLS to test robustness.
- Kernel Stress Tests: auditors perturb key kernel entries (KOPS) by ±0.05 to assess sensitivity of outcomes.

- Right-Threshold Challenges: auditors propose alternative rights thresholds (for example, stricter ecological integrity) and check if decisions reverse.
- PCC Minimalism Test: auditors attempt to reconstruct the decision using only the 5SPR and key parameters. If they cannot, the PCC is flagged for clarification.

Outcome reporting:

- Green: no material issues found.
- Yellow: minor issues, PCC revised with addendum.
- Red: major flaws, decision suspended pending re-analysis.

14. Validation, Falsification, and Empirical Program

14.1 Testable Hypotheses

MathGov is designed to be empirically evaluable. The framework suggests several hypotheses:

H1 (Rights coherence). Decisions that pass NCRC should produce fewer rights infringements in outcomes than decisions that fail NCRC, controlling for context.

H2 (Tail-risk effectiveness). Policies constrained by TRC should exhibit lower realized tail losses (frequency and severity) than comparable policies without TRC constraints.

H3 (Ripple predictiveness). After NCAR calibration, the sign of predicted impacts on key unions should match observed sign in at least approximately 70% of cases, improving over time.

H4 (UCI/HOI early warning). Positive HOI with negative UCI slope should predict downstream systemic failures better than baseline KPIs.

H5 (SGP parity). Entities that pass MIT-4 (legacy/deprecated; superseded by SGP 4.1.1) should demonstrate stable cross-union reasoning and learning patterns consistent with full-rights agents.

H6 (Cross-cultural invariance). The seven welfare dimensions should achieve configural, metric, and scalar invariance across diverse populations, or deviations should be explainable and correctable.

14.2 Validation Methods

An empirical program for MathGov can include:

Pre-registered pilots. Municipalities, organizations, or agencies use MathGov for selected decisions. Outcomes are tracked against pre-specified hypotheses and metrics.

Backtests. Historical decisions (policy changes, corporate strategies) are re-scored using MathGov to see whether NCRC/TRC would have flagged risks that did manifest.

Comparative studies. Compare decisions made under MathGov versus traditional cost-benefit analysis or ad hoc methods, controlling for context.

AI alignment experiments. Train RL agents or planning systems with and without MathGov constraints; compare emergent behaviors and safety properties.

Cross-cultural psychometrics. Collect data on welfare indicators across cultures; test dimension structure and invariance.

14.2.1 Backtesting Protocol (Retrospective Validation)

For retrospective validation, MathGov backtests must be reproducible and must avoid hindsight bias. The protocol is:

1. Select a historical decision with (a) a documented option set (or reconstructable alternatives), (b) a clear decision date, and (c) measurable outcomes 2-5 years later.
2. Reconstruct the decision-time information set: use only data, forecasts, constraints, and institutional context available at the time. Explicitly exclude post-decision facts.
3. Define the decision context (tier, time horizon, active unions and dimensions, applicability mask) and lock configuration versions (rights thresholds, TRC parameters, kernel profile, weights).
4. Specify affected subgroups for rights-covered cells; estimate subgroup impacts and compute worst-off subgroup impacts for NCRC.
5. Build direct impacts via the Canonical Impact Construction Algorithm (Section 3.2.7), including uncertainty intervals and provenance.
6. Apply propagation (Quick or Full) using the declared kernel profile; record stability checks and any fallbacks. (Tier 4 Pilot-Executable rev14.x: Quick only; see §13.8.2.)
7. Run the lexicographic cascade (NCRC → TRC → Containment → RLS → UCI/HOI) and record the model's ranked options and admissibility results.
8. Score predictive performance against realized outcomes using pre-declared outcome measures (by union and dimension where feasible), including subgroup outcomes for rights-relevant cells.
9. Report (a) sign accuracy, (b) calibration error (predicted vs observed magnitude bands), (c) whether NCRC and TRC would have flagged realized harms, and (d) whether UCI/HOI warnings preceded structural degradation.
10. Publish a "Backtest PCC" with full provenance and frozen configuration hashes so independent auditors can replay the backtest.

A backtest is invalid unless the reconstructed information set, configuration versions, and outcome measures are explicitly documented.

14.3 Open Science Commitments

MathGov adopts the following open-science principles to ensure transparency and reproducibility:

Pre-Registration: All pilot studies and validation tests are pre-registered on platforms like OSF or AsPredicted, with hypotheses, methods, and analysis plans locked before data collection.

Data Sharing: Anonymized PCC datasets, kernel profiles, and impact matrices are published under CC-BY licenses in standardized formats (JSON/CSV).

Code Availability: Reference implementations (Python/R) are open-source (MIT License) with versioned releases tied to validation phases.

Replication Challenges: Independent teams are invited to replicate key findings, with funding set aside for successful replications.

Error Bounties: Monetary bounties are offered for documented failures of MathGov in real-world tests (e.g., rights violations despite NCRC passage).

14.4 Falsification Criteria

MathGov makes strong empirical claims. It is falsified (in whole or part) if:

1. NCRC-passing decisions systematically produce worse rights outcomes than NCRC-failing decisions
2. TRC-constrained policies show no improvement in tail-risk metrics relative to unconstrained comparators, even after kernel updates
3. Ripple predictions remain systematically wrong ($< 60\%$ accuracy) across NCAR cycles
4. UCI and HOI fail to predict any meaningful aspect of system degradation in case studies
5. Dimension structure cannot be made invariant across a broad set of cultures

In such cases, the appropriate response is revision, not cosmetic adjustment. MathGov's legitimacy rests on its willingness to update or abandon components that fail empirically.

15. Applications and Illustrative Use Cases

15.1 AI Alignment and Governance

In AI, MathGov offers a constraint-first framework. Agents are trained only on admissible actions (NCRC and TRC satisfied), then optimize RLS within that set. Agents see not just a scalar reward but a structured 7×7 impact vector, enabling explainable, union-aware policies. When AI systems approach MI, SGP offers a principled route to grant rights parity.

Implementation architecture note: MathGov constraints can be implemented as hard-coded enforcement layers rather than as reward terms subject to gradient descent. This prevents the constraints from being "optimized away" during training. The NCRC and TRC filters operate as action-space projections that block inadmissible actions before reward evaluation occurs.

Example: Designing a resource-allocation AI for a hospital network. NCRC ensures no policy systematically denies basic care or violates bodily integrity. TRC prevents policies that risk catastrophic collapse of system capacity. RLS balances material efficiency, patient health, staff well-being, and long-term resilience. UCI prevents "efficiency" strategies that hollow out staff morale or trust.

15.2 Climate and Energy Policy

For climate mitigation and adaptation, NCRC can protect rights to life, health, and basic needs for vulnerable populations. TRC can encode existential climate scenarios. RLS can evaluate trade-offs between near-term economic impacts and long-term environmental and health outcomes. UCI can detect strategies that increase short-term prosperity but degrade structural cohesion.

MathGov provides a way to make explicit the union stack, including local communities, nations, humanity, and the biosphere, and to evaluate options such as fossil-fuel phase-out, adaptation investments, and energy subsidies with structured, auditable reasoning.

15.3 Organizational Strategy and Ethics

Organizations can use MathGov for major investments, product launches, restructuring decisions, safety and ethics reviews for new technologies, and internal policy changes affecting workers and communities.

Example: A multinational firm deciding whether to automate a labor-intensive process. NCRC ensures no policy violates workers' basic rights or ecological integrity. TRC addresses systemic risks. RLS evaluates multi-union impacts: efficiency gains, workforce displacement, community effects, environmental consequences. HDW allows stakeholders to participate in weighting within floors. NCAR ensures post-decision outcomes feed back into future strategy.

15.4 Personal and Small-Group Decisions

At micro scale, MathGov supports Tier 1 questions: Does this obviously violate someone's rights? Does it create catastrophic, irreversible risk? Among remaining options, which helps the most unions with the least harm? And Tier 2 approaches: small 3 × 3 or 4 × 4 matrices for family, project, or community decisions.

Even without full formalization, applying the lexicographic logic (rights, tails, ripples) can substantially improve decision quality and reduce avoidable harm.

16. Limitations, Risks, and Non-Targets

16.1 Meta-Fragility

The largest risk is meta-fragility: if core components are mis-set, MathGov can provide a veneer of legitimacy for harmful decisions, entrench biased floors or kernels as "scientific," and obscure responsibility behind complexity. Examples include rights floors calibrated too low for marginalized groups, kernels systematically underestimating certain environmental pathways, and HDW processes captured by powerful interests.

Mitigation: Floor Governance Charter (FGC) revision mechanisms; transparent publication of parameter settings; packaged critique and alternative models; robust validation and falsification efforts.

16.2 Misuse and Compliance Theater

MathGov could be misused as a rubber stamp (decisions made first, then parameters tweaked to justify them), a paper exercise (PCCs filled out perfunctorily), or a weapon (used to accuse opponents of being "anti-science" for questioning specific parameter choices).

Mitigation: Separate analysts from decision owners; random audits and public reporting of anomalies; training emphasizing that parameters are provisional; encouraging pluralism (multiple kernels, alternative weight profiles) instead of claiming a single infallible model.

16.3 Epistemic and Measurement Limitations

Not all welfare dimensions are equally easy to measure. Meaning and Agency may be harder to quantify than Material or Health. Data may be sparse or unreliable for certain unions or populations. Many ripple pathways will remain uncertain.

Mitigation: Use confidence scores and uncertainty penalties; prefer intervals over single-point estimates; drop or down-weight poorly measured dimensions explicitly, with justification; invest in better data and indicators over time.

16.4 Inter-Union Conflict

The framework assumes unions can be evaluated and weighted, but does not fully address cases where unions actively conflict (e.g., Organization vs Community, Polity A vs Polity B).

Mitigation: HDW weights represent a governance choice about relative priority. Persistent severe conflicts that cannot be resolved through weighting may require escalation to containing union governance (e.g., Polity-level conflicts escalated to CMIU-level coordination bodies). The Containment Principle provides partial protection by ensuring that gains to one union cannot come at the cost of degrading containing unions.

16.5 Computational Complexity

The framework's computational demands vary by tier:

1. Tier 1: Negligible, heuristic review (no quantitative simulation).

2. Tier 2: Low, quick calculable assessment using defaults (minimal scenario set if TRC is invoked).

3. Tier 3: Moderate, standard assessment with subgroup checks, uncertainty rules, and declared scenario library.

Organizations should assess computational requirements against decision timelines and ensure adequate infrastructure for intended tier of implementation.

16.6 Non-Target Domains

MathGov is not intended to replace interpersonal moral dialogue in intimate relationships, dictate individual aesthetic or spiritual preferences, or resolve all deep metaphysical disagreements. It is a framework for publicly accountable, multi-stakeholder decisions, not a totalizing substitute for personal conscience or cultural wisdom.

17. Conclusion: MathGov as a Universal Ethical Operating System

MathGov proposes a simple but demanding thesis: reality is relational, and we live in a network of unions whose fates are intertwined; sentient flourishing matters, and avoidable suffering should be reduced; rights and catastrophic risks are non-negotiable constraints that must be handled lexicographically before optimization; ripples and uncertainty matter, requiring explicit modeling of indirect effects while maintaining humility about what we know; and learning is essential, since no framework is final and systems must update themselves via evidence and reflection.

Operationally, MathGov provides a 49-cell welfare space that balances tractability and richness; a lexicographic cascade (NCRC → TRC → Containment → RLS → UCI) that forbids trading away rights for utility; an SGP and MI Plateau that ground rights in sentience and agency rather than species or raw intelligence; an HDW scheme that blends constitutional floors with democratic tuning; an explicit NCAR loop, PCC, and immutable audit structure to make decisions transparent and corrigible; and a validation and falsification program that invites empirical testing and revision.

Two meta-unions sit above the operational seven: the Cosmic Union, which will become practically relevant as humanity and other MI expand beyond Earth; and the AIU, representing the totality of existence and serving as a philosophical reminder of humility rather than a computational object.

MathGov is not a finished edifice but a scaffold: a coherent, mathematically specified system that can be implemented, tested, and improved. Its value will ultimately be measured by fewer rights violations, fewer catastrophic surprises, more resilient unions at every scale, and more transparent and accountable decisions.

If the framework succeeds, it will not be because it claims to be the final word on ethics, but because it gives humans and future MI better tools for asking the right questions, making structured trade-offs, learning from mistakes, and honoring the dignity of all who share the network of unions.

Conceptually, MathGov can be read as a governance-grade operationalization of established relational and systems thinking, distinguished by lexicographic rights and tail-risk constraints, explicit ripple propagation, and audit-ready decision artifacts. At its core, it is an operating system for alignment in a relational universe. It does not guarantee perfect outcomes, but it offers a principled way to seek better ones, systematically, transparently, and together.

Declarations

Funding

No packaged funding was received for this research.

Conflicts of Interest

The author declares no conflicts of interest relevant to this work.

Data Availability

This is a theoretical framework paper. No empirical datasets were generated or analyzed. Reference implementations and sample PCCs will be made available at mathgov.org following peer review.

Code Availability

Reference implementations in Python will be made available under MIT License at mathgov.org following peer review.

Author Contributions

The author conceived, designed, wrote, and revised the manuscript. Generative AI tools (OpenAI ChatGPT and Anthropic Claude) were used as writing and reasoning assistants during drafting, revision, and consistency checking. The author reviewed, verified, and edited all outputs and assumes full responsibility for the content, claims, and citations.

Ethics Approval

Not applicable. This work does not involve human subjects, animal research, or collection of personal data.

15.0 Packaging and Companion Artifacts (Normative)

This release is distributed as three coordinated documents: (1) MathGov Foundation Paper (this document), (2) MathGov Appendices Volume (companion), and (3) MathGov ProofPack v1.0 (companion artifact package). Together, these provide a complete Tier-4 Pilot-Executable specification.

Non-runnable by design. The Foundation Paper and Appendices are normative specifications, not executable code. They define required procedures, inputs, and validation rules. Implementations may be manual or software-assisted, but MUST conform to the specification.

ProofPack v1.0 (rev14, manifest-only). ProofPack provides the canonical, hash-bound scaffolding required for Tier-4 claims (schemas, manifests, canonicalization rules, and registries). It does not ship executable replay tooling in this release; implementations MUST maintain tool transparency (including tool hashes) sufficient for independent audit and replay against the ProofPack artifacts.

Binding by hash. For Tier-4 runs, the PCC MUST reference the exact Foundation Paper, Appendices, and ProofPack artifacts by SHA-256 (published in the Release Hashes document) and MUST reference all registry artifacts by SHA-256 as computed from the ProofPack canonicalization rules.

Completeness. A run that does not include a PCC with correct artifact hashes and configuration disclosure SHALL NOT be labeled Tier-4 Pilot-Executable, regardless of the quality of the decision analysis.

References

Amodei, D., Olah, C., Steinhardt, J., Christiano, P., Schulman, J., & Mané, D. (2016). Concrete problems in AI safety. *arXiv*. https://arxiv.org/abs/1606.06565

Arrow, K. J. (1963). *Social choice and individual values* (2nd ed.). Yale University Press.

Artzner, P., Delbaen, F., Eber, J.-M., & Heath, D. (1999). Coherent measures of risk. *Mathematical Finance, 9*(3), 203–228. https://doi.org/10.1111/1467-9965.00068

Aubin, J.-P. (1991). *Viability theory*. Birkhäuser.

Aubin, J.-P. (2009). *Viability theory* (2nd ed.). Springer.

Axelrod, R. (1984). *The evolution of cooperation*. Basic Books.

Bai, Y., et al. (2022). Constitutional AI: Harmlessness from AI feedback. arXiv. https://arxiv.org/abs/2212.08073

Bai, Y., Jones, A., Ndousse, K., Askell, A., Chen, A., DasSarma, N., Drain, D., Fort, S., Ganguli, D., Henighan, T., Joseph, N., Kadavath, S., Kernion, J., Conerly, T., El-Showk, S., Elhage, N., Hatfield-Dodds, Z., Hernandez, D., Hume, T., ... Kaplan, J. (2022b). Training a helpful and harmless assistant with reinforcement learning from human feedback. *arXiv*. https://arxiv.org/abs/2204.05862

Barabási, A.-L., & Albert, R. (1999). Emergence of scaling in random networks. *Science, 286*(5439), 509–512. https://doi.org/10.1126/science.286.5439.509

Becker, G. S. (1981). *A treatise on the family*. Harvard University Press.

Beer, S. (1972). *Brain of the firm*. Allen Lane.

Beer, S. (1979). *The heart of enterprise*. Wiley.

Belton, V., & Stewart, T. J. (2002). *Multiple criteria decision analysis: An integrated approach*. Kluwer Academic Publishers.

Bostrom, N. (2014). *Superintelligence: Paths, dangers, strategies*. Oxford University Press.

Bowlby, J. (1988). *A secure base: Parent-child attachment and healthy human development*. Basic Books.

Christakis, N. A., & Fowler, J. H. (2009). *Connected: The surprising power of our social networks and how they shape our lives*. Little, Brown and Company.

Christian, B. (2020). *The alignment problem: Machine learning and human values*. W. W. Norton & Company.

Cooke, R. M. (1991). *Experts in uncertainty: Opinion and subjective probability in science*. Oxford University Press.

Cryan, J. F., & Dinan, T. G. (2012). Mind-altering microorganisms: The impact of the gut microbiota on brain and behaviour. *Nature Reviews Neuroscience, 13*(10), 701–712. https://doi.org/10.1038/nrn3346

Dehaene, S. (2014). *Consciousness and the brain: Deciphering how the brain codes our thoughts*. Viking.

Dolan, P., Peasgood, T., & White, M. (2008). Do we really know what makes us happy? A review of the economic literature on the factors associated with subjective well-being. *Journal of Economic Psychology, 29*(1), 94–122. https://doi.org/10.1016/j.joep.2007.09.001

Dryzek, J. S. (2000). *Deliberative democracy and beyond: Liberals, critics, contestations*. Oxford University Press.

Dunbar, R. I. M. (1992). Neocortex size as a constraint on group size in primates. *Journal of Human Evolution, 22*(6), 469–493. https://doi.org/10.1016/0047-2484(92)90081-J

Dunbar, R. I. M. (1993). Coevolution of neocortical size, group size and language in humans. *Behavioral and Brain Sciences, 16*(4), 681–735. https://doi.org/10.1017/S0140525X00032325

Durand, M., & Boarini, R. (2016). *How's Life 2015: Measuring Well-being*. OECD Publishing. https://doi.org/10.1787/how_life-2015-en

Estes, J. A., et al. (2011). Trophic downgrading of planet Earth. Science, 333(6040), 301–306. https://doi.org/10.1126/science.1205106

Fishkin, J. S. (2009). *When the people speak: Deliberative democracy and public consultation*. Oxford University Press.

Food and Agriculture Organization of the United Nations. (2021). *Voices of the hungry: The Food Insecurity Experience Scale*. FAO. https://www.fao.org/in-action/voices-of-the-hungry/fies/en/

Freedom House. (2024). *Freedom in the World 2024: The annual survey of political rights and civil liberties*. Freedom House. https://freedomhouse.org/report/freedom-world

Fung, A. (2006). Varieties of participation in complex governance. *Public Administration Review, 66*(Suppl. 1), 66–75. https://doi.org/10.1111/j.1540-6210.2006.00667.x

Goodhart, C. A. E. (1984). Problems of monetary management: The UK experience. In C. A. E. Goodhart (Ed.), *Monetary theory and practice: The UK experience* (pp. 91–121). Macmillan.

Gray, M. W. (2012). Mitochondrial evolution. *Cold Spring Harbor Perspectives in Biology, 4*(9), a011403. https://doi.org/10.1101/cshperspect.a011403

Habermas, J. (1996). *Between facts and norms: Contributions to a discourse theory of law and democracy*. MIT Press.

Holt-Lunstad, J., Smith, T. B., Baker, M., Harris, T., & Stephenson, D. (2015). Loneliness and social isolation as risk factors for mortality: A meta-analytic review. *Perspectives on Psychological Science, 10*(2), 227–237. https://doi.org/10.1177/1745691614568352

Hume, D. (1978). *A treatise of human nature* (L. A. Selby-Bigge & P. H. Nidditch, Eds.). Oxford University Press. (Original work published 1739)

Intergovernmental Panel on Climate Change. (2023). *Climate change 2023: Synthesis report. Contribution of Working Groups I, II and III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change* (Core Writing Team, H. Lee, & J. Romero, Eds.). IPCC. https://doi.org/10.59327/IPCC/AR6-9789291691647

International Organization for Standardization, & International Electrotechnical Commission. (2023). *ISO/IEC 23894:2023 Artificial intelligence: Guidance on risk management*. ISO.

Jones, K. E., Patel, N. G., Levy, M. A., Storeygard, A., Balk, D., Gittleman, J. L., & Daszak, P. (2008). Global trends in emerging infectious diseases. *Nature, 451*(7181), 990–993. https://doi.org/10.1038/nature06536

Keeney, R. L., & Raiffa, H. (1976). *Decisions with multiple objectives: Preferences and value tradeoffs*. Wiley.

Lenton, T. M., Held, H., Kriegler, E., Hall, J. W., Lucht, W., Rahmstorf, S., & Schellnhuber, H. J. (2008). Tipping elements in the Earth's climate system. *Proceedings of the National Academy of Sciences, 105*(6), 1786–1793. https://doi.org/10.1073/pnas.0705414105

Leontief, W. (1986). *Input-output economics* (2nd ed.). Oxford University Press.

Manheim, D., & Garrabrant, S. (2018). Categorizing variants of Goodhart's law. *arXiv*. https://arxiv.org/abs/1803.04585

Mansuri, G., & Rao, V. (2013). *Localizing development: Does participation work?* World Bank. https://doi.org/10.1596/978-0-8213-8256-1

March, J. G., & Simon, H. A. (1958). *Organizations*. Wiley.

Margulis, L. (1998). *Symbiotic planet: A new look at evolution*. Basic Books.

Marmot, M. (2010). *Fair society, healthy lives: The Marmot Review. Strategic review of health inequalities in England post-2010*. The Marmot Review. https://www.instituteofhealthequity.org/resources-reports/fair-society-healthy-lives-the-marmot-review

Max-Neef, M. A. (1991). *Human scale development: Conception, application and further reflections*. Apex Press.

Meadows, D. H. (2008). *Thinking in systems: A primer*. Chelsea Green Publishing.

National Institute of Standards and Technology. (2023). *Artificial Intelligence Risk Management Framework (AI RMF 1.0)* (NIST AI 100-1). https://doi.org/10.6028/NIST.AI.100-1

Newman, M. E. J. (2010). *Networks: An introduction*. Oxford University Press.

Nordhaus, W. D. (2017). Revisiting the social cost of carbon. *Proceedings of the National Academy of Sciences, 114*(7), 1518–1523. https://doi.org/10.1073/pnas.1609244114

North, D. C. (1990). *Institutions, institutional change and economic performance*. Cambridge University Press.

Nowak, M. A. (2006). Five rules for the evolution of cooperation. *Science, 314*(5805), 1560–1563. https://doi.org/10.1126/science.1133755

Nussbaum, M. C. (2011). *Creating capabilities: The human development approach*. Harvard University Press.

Odum, E. P. (1971). *Fundamentals of ecology* (3rd ed.). W. B. Saunders.

Office of the High Commissioner for Human Rights. (2012). *Human rights indicators: A guide to measurement and implementation*. United Nations. https://www.ohchr.org/Documents/Publications/Human_rights_indicators_en.pdf

Ord, T. (2020). *The precipice: Existential risk and the future of humanity*. Hachette Books.

Organisation for Economic Co-operation and Development. (2019). *Recommendation of the Council on Artificial Intelligence* (OECD/LEGAL/0449). OECD. https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449

Popper, K. (1959). *The logic of scientific discovery*. Hutchinson.

Putnam, R. D. (2000). *Bowling alone: The collapse and revival of American community*. Simon & Schuster.

Rawls, J. (1971). *A theory of justice*. Harvard University Press.

Raworth, K. (2017). *Doughnut economics: Seven ways to think like a 21st-century economist*. Chelsea Green Publishing.

Rockafellar, R. T., & Uryasev, S. (2000). Optimization of conditional value-at-risk. *Journal of Risk, 2*(3), 21–42. https://doi.org/10.21314/JOR.2000.038

Rockström, J., et al. (2009). A safe operating space for humanity. Nature, 461, 472–475. https://doi.org/10.1038/461472a

Rockström, J., et al. (2023). Safe and just Earth system boundaries. Nature, 619, 102–111. https://doi.org/10.1038/s41586-023-06083-8

Russell, S. (2019). *Human compatible: Artificial intelligence and the problem of control*. Viking.

Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist, 55*(1), 68–78. https://doi.org/10.1037/0003-066X.55.1.68

Sen, A. (1999). *Development as freedom*. Knopf.

Sen, A. (2009). *The idea of justice*. Harvard University Press.

Simon, H. A. (1962). The architecture of complexity. *Proceedings of the American Philosophical Society, 106*(6), 467–482.

Sphere Association. (2018). *The Sphere handbook: Humanitarian charter and minimum standards in humanitarian response* (4th ed.). Sphere Association. https://spherestandards.org/handbook/

Steffen, W., Richardson, K., Rockström, J., Cornell, S. E., Fetzer, I., Bennett, E. M., Biggs, R., Carpenter, S. R., de Vries, W., de Wit, C. A., Folke, C., Gerten, D., Heinke, J., Mace, G. M., Persson, L. M., Ramanathan, V., Reyers, B., & Sörlin, S. (2015). Planetary boundaries: Guiding human development on a changing planet. *Science, 347*(6223), 1259855. https://doi.org/10.1126/science.1259855

Stiglitz, J., Fitoussi, J. P., & Durand, M. (2018). *Beyond GDP: Measuring what counts for economic and social performance*. OECD Publishing. https://doi.org/10.1787/9789264307292-en

Taleb, N. N. (2012). *Antifragile: Things that gain from disorder*. Random House.

Tononi, G. (2008). Consciousness as integrated information: A provisional manifesto. *Biological Bulletin, 215*(3), 216–242. https://doi.org/10.2307/25470707

Tsemberis, S., Gulcur, L., & Nakae, M. (2004). Housing First, consumer choice, and harm reduction for homeless individuals with a dual diagnosis. *American Journal of Public Health, 94*(4), 651–656. https://doi.org/10.2105/AJPH.94.4.651

United Nations High Commissioner for Refugees. (2024). *Emergency handbook*. UNHCR. https://emergency.unhcr.org/

Weber, M. (1978). *Economy and society* (G. Roth & C. Wittich, Eds.). University of California Press. (Original work published 1922)

World Health Organization. (2022). *World health statistics 2022: Monitoring health for the SDGs*. WHO. https://www.who.int/publications/i/item/9789240051157

World Justice Project. (2023). *Rule of Law Index 2023*. World Justice Project. https://worldjusticeproject.org/rule-of-law-index/

Young, I. M. (2000). *Inclusion and democracy*. Oxford University Press.

Zilber-Rosenberg, I., & Rosenberg, E. (2008). Role of microorganisms in the evolution of animals and plants: The hologenome theory of evolution. *FEMS Microbiology Reviews, 32*(5), 723–735. https://doi.org/10.1111/j.1574-6976.2008.00123.x