

Comparability across Heterogeneous Generative Systems

A Preliminary Inquiry into Cross-Relational Invariance

Wanhong HUANG
huangwanhong@serendip.ngo

Abstract

Similar observable patterns across heterogeneous systems can conceal differences in admissible transformations, resources, histories, and interpretation. This paper examines the conditions under which comparative inferences remain justified across such differences. It develops a task-relative account of comparative warrant organized around declared objects, partial relational mappings, selected invariants, conditioning regimes, available evidence, and criteria for revision. Conditional mathematical results establish properties of exact relational quotients, composition and structural preservation, resource-sensitive feasibility, history-sensitive continuation, partial identification, reciprocal translation, and ranking stability. Their proofs use established mathematical constructions and warrant physical or interpretive conclusions only through additional evidence. Comparative designs involving biological regulation, ecological networks, computational processes, and linguistic annotation distinguish documented findings from illustrative models and prospective empirical tests. The analysis further addresses the authority to select comparative categories and proposes a transparent warrant record for evaluating inference, uncertainty, and opportunities for correction. Comparison against established specialist workflows remains an open empirical task. The contribution is a coordinated framework for assessing the scope, preservation requirements, and revisability of cross-system claims, under ontological suspension.

Keywords: heterogeneous generative systems; cross-relational invariance; partial correspondence; comparative warrant; epistemic translation

Discussion Paper Note

This paper develops an exploratory philosophical and methodological research programme on comparability among heterogeneous generative-relational systems. Its principal object is the warrant for translating or comparing relational patterns, transformations, and generative possibilities across systems whose descriptions, material conditions, and historical trajectories may differ. The research is related to earlier inquiries into generative order, political-economic relations among heterogeneous subjects, and evolutionary norm formation; it pursues a distinct question concerning cross-system correspondence and the possible preservation of relational structure.

The discussion draws selectively on philosophy of scientific representation, debates on incommensurability, structural realism, category theory, model theory, and research on measurement invariance. These are established intellectual traditions. Terms such as *cross-relational invariance* and *comparative warrant* designate provisional organizational concepts within the present inquiry. The initial formal descriptions and illustrative cases specify the scope of comparison; their broader evaluative and empirical applications remain proposals for investigation.

Version 1.5.1 presents twelve substantive sections spanning the comparative object, partial mappings, selected invariants, joint generative feasibility, historical provenance, underidentification, reciprocal epistemic translation, comparative cases, and the institutional authority to select and revise comparison categories, and a critical methodology for evaluating comparative warrants against specialist alternatives. The Conclusion synthesizes the conditional results, methodological proposals, and outstanding empirical tests. The manuscript reports conditional mathematical claims and illustrative cases without completed empirical interventions. The author invites criticism and comparative research, especially concerning failed translations, changing representations, category design, and unequal opportunities for correction.

Hypothetical examples involving unfamiliar biological, artificial, or extraterrestrial systems are analytical constructions; they imply no factual conclusion about the existence, experience, or status of any particular entity.

Responsible Use and Rights Reservation

This notice requests responsible scholarly use and remains separate from the copyright licence below. The paper welcomes independent criticism, careful reinterpretation, and scholarly comparison. Applications to scientific classification, governance, artificial intelligence, education, or cross-cultural interpretation require context-sensitive evaluation and suitable engagement with affected participants. A proposed correspondence between systems may omit relevant histories, capabilities, or stakes, and its use should therefore remain contestable and revisable.

Except where otherwise indicated, copyright © 2026 Wanhong Huang. Original materials in this source package are supplied under the Creative Commons Attribution–NonCommercial 4.0 International licence (CC BY-NC 4.0). Third-party material remains under its respective rights. Licence text: <https://creativecommons.org/licenses/by-nc/4.0/>. This advisory statement does not modify the licence or introduce additional contractual conditions.

Notices

Research scope. Independent unreviewed preprint, version 1.5.1. All twelve substantive sections are included. The manuscript offers provisional conceptual arguments; it includes no original empirical measurements or physical validation.

Licence. Unless otherwise specified, copyright © 2026 Wanhong Huang. Original content is licensed under Creative Commons Attribution–NonCommercial 4.0 International (CC BY-NC 4.0), subject to applicable law. The responsible-use statement is advisory and does not diminish rights granted by that licence.

AI-assisted preparation. OpenAI’s ChatGPT assisted with research dialogue, source discovery, English drafting, and LaTeX preparation. The human author directed the research programme and remains responsible for interpretation, bibliographic accuracy, further review, and dissemination. AI assistance does not constitute independent peer review.

Independence. The inquiry is independently directed. Citation of a scholar, institution, or programme implies neither endorsement nor collaboration.

Suggested citation. Huang, Wanhong. “Comparability across Heterogeneous Generative Systems: A Preliminary Inquiry into Cross-Relational Invariance.” Unreviewed preprint, v1.5.1, 2026.

1 Introduction

Comparative inquiry frequently proceeds by identifying relations that appear sufficiently similar to support an inference across distinct systems. A recurrent transformation in two physical models may suggest dynamical correspondence; an analogous pattern of production in biological and artificial systems may motivate a comparison of regenerative capacities; and similarities in communication may prompt interpretations of another participant's commitments or expectations. Each comparison depends on a decision about which features matter, which differences remain acceptable, and what inferential work the proposed correspondence can sustain. The present paper investigates these decisions in the context of heterogeneous generative-relational systems. Its purpose is to formulate a preliminary account of the conditions under which cross-system comparison becomes intelligible, informative, and critically revisable.

A generative-relational system is understood here as a candidate representation of configurations, relational conditions, and admissible practices or transformations through which further configurations may be realized. The expression is intentionally broad. Its extension depends on an explicit model and research purpose; a candidate system may describe a chemical reaction network, an evolving ecological arrangement, a computational architecture, or a historically situated form of social cooperation. Such uses share an analytical vocabulary while retaining potentially different ontological commitments and evidential requirements. Generativity describes admissible transformations under stated conditions. Relationality directs attention to configurations and dependencies constituted or conditioned by interactions. Neither term establishes, by itself, the metaphysical nature of the systems to which it is applied.

The investigation is motivated by a familiar ambiguity. Two systems can display an isomorphic diagram of observable transitions while relying on different material resources, internal constraints, or mechanisms of persistence. Conversely, systems organized by very different descriptions can generate comparable consequences under an agreed observation procedure. Comparing their diagrams may therefore be sound for a limited task while remaining inadequate for a broader inference about causal mechanism, practical substitutability, or normative significance. The relevant issue is the scope of a correspondence: which relations are being preserved, within which observational domain, and under what conditions does preservation authorize conclusions outside that domain?

Existing philosophies of science have developed closely related questions. Incommensurability debates, associated with Kuhn and Feyerabend, examine conceptual and methodological differences between scientific traditions. The contemporary analysis makes clear that incommensurability need not entail an absolute inability to compare theories; local conceptual incompatibilities can coexist with shared empirical predictions or partial forms of appraisal [24]. The present inquiry inherits that caution while extending its immediate object from scientific theories to candidate generative systems and their relations. Its proposed extension requires justification in each application. Scientific theories are articulated representations subject to disciplinary standards; biological, social, and artificial systems may have causal and material properties beyond the theories used to investigate them. A comparison of descriptions is accordingly distinguished from an assertion about correspondence between the systems represented.

Davidson's critique of the idea of a conceptual scheme provides a further challenge to any account that treats complete conceptual separation as straightforwardly intelligible [7]. The relevance of this critique is methodological. If the ascription of radically different schemes already relies on interpretation, then an inquiry into heterogeneity should specify how the alleged difference was recognized. Equally, recognizing a local interpretive success gives limited grounds for declaring a universal semantic translation. These considerations support the use of graded, task-specific and revisable comparative claims. A failure to translate one predicate or predictive quantity can coexist with successful coordination in another domain; a locally effective translation can also leave crucial practices unrepresented.

The philosophy of scientific representation likewise cautions against identifying structural resemblance with successful representation. Formal isomorphism describes a mathematical relation between structures, while representational success depends on the target, purpose, interpretation, and standards of adequacy [13]. For G4 this distinction becomes particularly important when a diagram, graph, category,

field, or probabilistic model is employed as the vehicle of comparison. The fact that two descriptions can be linked by a structure-preserving map establishes a property of those descriptions. A justified claim concerning the target systems requires separate evidence for the descriptive adequacy of each model and for the significance of the properties preserved by the proposed map.

A mathematically precise tradition is available for studying such maps. Category theory distinguishes objects and morphisms within a specified category, while functors preserve identities and composition between categories [20]. Model theory examines structures, interpretations, and definability in a different formal setting [16]. These tools are valuable precisely because their preservation conditions can be stated. Yet an unrestricted transfer from a functorial correspondence to claims of physical equivalence would overreach: a functor preserves the structure that has been formalized and need not preserve resources, empirical probabilities, institutional histories, or ethical standing omitted from the categorical presentation. The main text will therefore treat category-theoretic constructions as optional, task-bound representations. Where a simpler relation, partial function, or span suffices, the argument should use that simpler formulation.

An additional conceptual resource comes from research on measurement invariance. Comparative measurement across groups requires examination of whether indicators function in sufficiently equivalent ways; some comparisons may remain meaningful under carefully assessed partial invariance [27]. The precise psychometric results do not automatically extend to other domains. Their methodological lesson is nonetheless instructive: the use of a common label or numerical scale does not alone guarantee that measured quantities represent the same underlying property. In a cross-system generative inquiry, even the choice of a purportedly shared outcome can hide differences in how an outcome is produced, recognized, or maintained. This motivates explicit specification of observable domains, identification assumptions, and the consequences of failing those assumptions.

The present paper proposes to distinguish several questions that are often compressed into one general assertion of similarity. Descriptive correspondence concerns whether selected configurations or relations can be associated within a declared vocabulary. Structural preservation concerns whether a proposed mapping retains specified formal operations, connections, or invariants. Generative correspondence asks whether related configurations support matched admissible transformations, possibly under different conditioning regimes. Evidential correspondence concerns whether the resulting comparisons survive observations, interventions, or other forms of warranted assessment. Normative comparability concerns whether the proposed common terms of evaluation respect the relevant differences in value, standing, vulnerability, and authority. These questions have different criteria, and success at one level supplies an input to further inquiry rather than a comprehensive answer.

The relations among these levels are neither uniform nor inevitably hierarchical. A physical model may admit a formal structural correspondence that lacks stable empirical support. A comparative observational study may recover a robust shared regularity despite incomplete knowledge of underlying mechanisms. An institutional encounter may display reliable coordination although participants assign different meanings to their practices. A normative agreement may provide workable common standards for a particular decision while preserving disagreements about historical responsibility and legitimate authority. Such possibilities encourage an analysis of the specific inferential passage being attempted and of the conditions under which that passage is warranted.

For preliminary orientation, let $\mathfrak{S}_a$ and $\mathfrak{S}_b$ denote two candidate generative-system descriptions, each with its own admissible configurations and transformations. A cross-system correspondence Φ may associate only selected parts of these descriptions. The central task is to determine which properties, collected provisionally as a family $\mathcal{I}$, are preserved by Φ within a declared domain $\mathcal{D}$ and under an evidential regime $\mathcal{E}$. This notation is an organizational aid, and later sections will specify its alternative formal realizations. The expression *cross-relational invariance* designates the conditional persistence of selected relational properties across such a correspondence. Its use should remain limited to properties whose preservation has been defined and supported; it neither posits an invariant essence of all generative systems nor implies the existence of a privileged universal language.

The distinction between preservation and legitimacy is especially significant in encounters involving

radically different backgrounds. A shared measure may be chosen because it is accessible to a powerful observer, institution, or platform, while a less powerful participant's generative conditions remain outside the comparison. A mapping may be mathematically correct on the domain selected by an investigator while representing that investigator's priorities. Comparative methodology therefore carries an additional reflective obligation: to explain how its domain was selected, whose distinctions it suppresses, and how the scope or terms of comparison may be challenged. This obligation follows from the fallibility and situatedness of inquiry; it is developed without treating any single normative framework as settled in advance.

The paper's relationship to the companion research programme is specific. The inquiry into the ontological status of generative order (G1) examines derivation of precedence, representational equivalence, and conditions for physical interpretation within candidate landscapes. The political-economic inquiry (G2) investigates heterogeneous production, reproduction, power, value circulation, and institutional contestability. The evolutionary norm-formation inquiry (G3) concerns how provisional norms emerge and change through interaction among heterogeneous participants. G4 asks a more general epistemological and methodological question: when relations developed in one system, research tradition, or institutional setting may be compared with relations in another, and what remains valid after translation. These inquiries can inform each other without presuming that a solution in one settles the problems of the others.

The comparative cases use biological regulation, ecological networks, computational systems, and linguistic annotation to distinguish warranted inferences from diagrammatic resemblance. Their evidential basis and limits are assessed in Section 9. Hypothetical encounters with unfamiliar systems serve as bounded tests of the assumptions used to construct a comparison; they provide analytical contrasts rather than evidence about actual unfamiliar beings.

The proposed account distinguishes successful partial comparison, failed mapping, indeterminate correspondence, and contested comparative standards. Several locally reliable mappings may coexist with unresolved remainders. Explicit preservation criteria support precise inferences where their assumptions are met; documented limitations provide grounds for revising the mapping or the intended use of its results.

The twelve-section analysis begins with comparative objects, partial relational correspondences, selected invariants, generative compatibility, and historical provenance. It then addresses observational reconstruction, reciprocal epistemic translation, natural and artificial comparative cases, and authority over comparative categories. The final analytical section specifies evaluation protocols and revision criteria; the Conclusion integrates the formal, interpretive, and institutional findings. All twelve substantive sections are presented in this version, with empirical evaluation of the proposed coordination method remaining a future research task.

G4 accordingly proposes a task-relative protocol for examining cross-system comparative warrant through declared representations, partial mappings, preservation criteria, evidence, and revision opportunities. Its additional usefulness will require testing against established specialist workflows, including whether the protocol helps identify invalid analogies, interpret heterogeneous evidence, or improve accountable comparative decisions. Ontological suspension remains appropriate: a scientifically useful invariant can support a specific inference while leaving the ultimate relation among the represented systems open.

2 The Comparative Object in Heterogeneous Systems

A comparative inquiry into heterogeneous generative systems requires an explicit account of what is being compared. The present section establishes that account before proposing transformations between systems. Its objective is to distinguish the target of inquiry from its description, the configuration of a system from the conditions under which transformations are admissible, and an observed regularity from the history and evidential practices through which the regularity is known. The discussion proceeds through conceptual analysis, a deliberately economical formal representation, and two schematic

comparisons. It closes by specifying the information that a defensible comparison should carry into the partial-mapping analysis of Section 3. The resulting account describes a *comparative object*: a situated relation between selected descriptions and questions, accompanied by the evidence and qualifications needed to assess an inference.

2.1 Targets of Inquiry and Descriptive Commitments

A system encountered through research can rarely be taken up in its entirety. An investigator selects a domain of observation, identifies candidate configurations and relations, introduces a vocabulary, and determines the kinds of differences that will be registered. A chemical network may be described by molecular species and reaction channels; an ecological system by organisms, flows, and environmental constraints; a computational process by states and transition rules; and a communicative practice by participants, utterances, interpretations, and responses. Each representation supplies access to some questions and leaves others outside its declared scope. A comparison between systems therefore begins with a pair of interpreted descriptions whose targets, construction procedures, and intended uses are specified.

We distinguish a *target* X_a from a *candidate description* M_a of that target. The symbol X_a indexes the part of the world, practice, or stipulated hypothetical scenario under investigation. It does not presuppose that the target has a unique decomposition into objects or an observer-independent list of properties. The description M_a contains distinctions and relations that an inquiry proposes as relevant to a question. For an abstract mathematical example, the target may itself be a stipulated structure and the adequacy relation can be exact. For an empirical study, the link between X_a and M_a requires observational, experimental, interpretive, or historical support. This distinction is consistent with accounts of scientific representation in which formal structure alone leaves questions of interpretation and representational adequacy unsettled [13].

The preliminary use of a shared description type across two targets is a methodological choice. A pair of systems can be described using similar data structures while calling for different measurement practices, different physical mechanisms, or different assessments of error. Conversely, distinct descriptive languages can be adequate to a shared task if a suitable interpretation relates their outputs. For that reason, the comparative object includes a statement of the research question q and an account of what would count as a successful answer. A description developed for forecasting one observable may have little bearing on causal explanation, long-horizon feasibility, or the justification of institutional treatment. It is helpful to say that descriptions are *adequate for particular inferences* rather than adequate without qualification.

The distinction also prevents a common slide from descriptive similarity to similarity of the represented targets. Suppose two models have isomorphic finite transition graphs. This mathematical fact establishes a correspondence between their selected state-transition structures. If one model records reusable symbolic tokens and the other represents consumable material stocks, the graph isomorphism omits a difference relevant to the feasibility of repeated production. The selected graph may be fully appropriate for a restricted enumeration task, while a comparison of sustainable productive capacity requires a richer description. The material or institutional significance of any omitted feature must be established within the relevant domain; it cannot be inferred from the vocabulary of heterogeneity alone.

2.2 A Minimal Representation of a Candidate Generative System

For organizational purposes, a candidate system description may be represented by the tuple in Equation 1. The construction is proposed as a vocabulary for recording assumptions and comparing research tasks. It asserts neither a universal ontology of systems nor an empirical model valid across disciplines.

$$M_a = (D_a, C_a, K_a, \mathcal{G}_a, \mathcal{O}_a, \mathcal{H}_a). \quad (1)$$

Definition 2.1 (Interpreted candidate-system description). In Equation 1, D_a is a set of admissible descriptive expressions and interpretation rules; C_a is the set of candidate configurations represented within that description; and K_a is a family of specified *regimes* of environmental, material, institutional, or interpretive conditions. The component $\mathcal{G}_a$ is a family of admissible generative relations, potentially dependent on $\kappa \in K_a$ and involving several jointly required configurations or resources. The component $\mathcal{O}_a$ records observation or interpretation interfaces, including their domains and resolution; and $\mathcal{H}_a$ records the provenance, histories, or transformations relevant to the research question. The tuple is *interpreted* only together with an explicit account of how its elements are associated with an intended target and assessed against available evidence.

The tuple permits heterogeneous formal realizations. The configurations C_a might be sets of graph-labelled states, probability distributions, material inventories, or documented positions within a practice. A regime κ might fix reaction conditions, computational resources, a social institution, or the circumstances under which a linguistic expression is interpreted. Generative relations may be functions, stochastic kernels, partial relations, or constraints on jointly feasible practices. In a setting that depends on continuous fields or unbounded state spaces, these components need corresponding mathematical structures and regularity assumptions. The tuple is accordingly a schematic signature, not a claim that all relevant phenomena admit tractable finite representations.

A generic relation $G_{a,\kappa}^{(m)} \subseteq C_a^m \times C_a$, where $m \geq 1$, can record a permitted output under m jointly specified premises. This notation represents a possibility relative to a regime and a declared description. It gives no general license to infer that the premises are independently available, reusable, physically realized, or normatively desirable. Applications may require resource inventories, quantitative costs, probabilistic transition laws, or contextual constraints beyond a simple relation. Section 3 and subsequent formal discussions will specify preservation requirements for selected kinds of relations; the present section only identifies the object to which those requirements must refer.

The observational component $\mathcal{O}_a$ deserves similar care. An observation map records the results accessible under a particular procedure. Several underlying configurations may share the same observed output, and the same configuration may yield different readings under alternative instruments or interpretive conventions. The associated errors and ambiguities become part of the comparative object when they affect the proposed inference. A common observable is therefore a promising *interface for inquiry*, with its evidential significance dependent on what each observational procedure establishes.

The historical component $\mathcal{H}_a$ is intentionally broad. It can include records of construction, constraints inherited from earlier transformations, versions of a model, chains of evidence, or documented organizational commitments. Its role is to preserve dependencies that could be hidden by a static description. The expression *history* here designates a structured account of relations among events, transformations, or conditions; it carries no commitment to an ontologically privileged present or to a universal chronological organization. The research question determines which of these relations materially affect comparison.

2.3 Regimes, Admissibility, and the Locality of Generative Claims

A purported generative capacity is incomplete when its conditions of operation are unspecified. A reaction may occur only at particular concentrations, temperatures, or available energy; a computational process may depend on memory and platform services; and a form of cooperation may depend on rules of participation, maintenance arrangements, or reciprocal expectations. Within each domain, the conditions of admissibility determine which transformations can count as feasible. Comparisons that omit these conditions may conflate a rule’s formal description with a materially or institutionally realized possibility.

There are at least three analytically distinct kinds of conditioning. *Constitutive conditions* define the domain of the description and the admissibility of its configurations. *Operational conditions* govern which transformations can be executed or sustained within an admissible domain. *Evidential conditions* determine how the investigator can know whether a transformation is available, performed, or recurrent.

A mathematical automaton may have a completely specified operational relation by stipulation. A biological or institutional description may require empirical inquiry before the same kind of assertion is warranted. While the categories can overlap in an application, declaring them separately makes the basis of a comparison more inspectable.

In particular, the equation of *reachability* with *capacity* depends on the selected meaning of the latter. A configuration might be reachable once under a one-off resource expenditure, reachable repeatedly after replenishment, or reachable only under a regime that alters the participant's future options. The first statement concerns the existence of a derivation. The second and third require additional information about resources, continuation, and changing conditions. An observer interested in comparative adaptability may need to examine the family of regimes under which an outcome remains feasible, including the costs of switching regimes. An observer interested in practical reliance may need to examine whether such regimes can be maintained without imposing unexamined burdens on other participants.

A condition-sensitive claim should record at least the starting descriptions, the admissible regime, the relevant input constraints, the output domain, and any horizon of operation. These elements can be supplied qualitatively where quantitative characterization is unwarranted. Their importance lies in preventing one interpretation of capacity from silently replacing another. Comparability remains local even where a common vocabulary makes broader extrapolation tempting.

Differences in conditions also matter when a generative system is represented at several scales. A coarse model may supply effective transition rules that summarize microscopic activity, while a finer model records materials, dependencies, and mechanisms omitted at the coarse scale. An effective description can be adequate to a particular observable task without furnishing a faithful reconstruction of each microscopic pathway. Whether a proposed cross-scale relation preserves the generative property of interest must be examined rather than assumed. G1 investigates several mathematical versions of this issue within one generative landscape; G4 addresses the additional challenge of interpreting purportedly corresponding properties across separately constructed descriptions.

2.4 Units, Observables, and the Meaning of a Common Measure

Comparative judgments frequently become persuasive through apparently shared units or indices. Two systems may be assigned a throughput, persistence rate, accuracy score, production count, or coordination measure. The use of a common unit facilitates a numerical comparison, provided that the measurement procedures, denominators, and target attributes are aligned sufficiently for the intended question. Such alignment has several dimensions: the terms used to designate an attribute, the operational procedure for generating a value, the relationship between that value and the attribute, and the permissible interpretation of differences across systems.

A simple output count illustrates the issue. In one arrangement, a produced token may be freely copied; in another, an analogous physical product may require scarce material and maintenance. Equal counts can describe equivalent performance on a narrowly defined output task while yielding different conclusions about the resources, causal routes, and future capabilities that sustain production. A numerical alignment becomes informative when the relevant omitted dimensions are either immaterial to the question or independently assessed. The research should make that choice explicit.

Measurement invariance research provides a disciplined example of the care required in interpreting comparable scores. In psychometrics, comparisons across populations depend on assumptions concerning the relationships among measured responses, latent constructs, and group membership; those assumptions can be investigated rather than read directly from similar numerical values [27]. G4 borrows the methodological question of alignment, not a general theorem that psychometric criteria apply unchanged to ecological, computational, or physical systems. Each domain must identify the conditions under which a shared measurement is an adequate proxy for the comparative property under study.

Shared observables may also be created through an interpretive transformation. Different instruments can report quantities that become comparable after calibration; different institutional categories can be aligned by a negotiated coding convention; distinct semantic analyses can support common predictions

under specified contexts. These operations carry information about the origin and limits of comparability. A transformation that preserves one statistic may change uncertainty, erase distinctions, or reassign the meaning of an outcome. For that reason, the claim *the systems have the same measured property* is strongest when accompanied by a statement of which property, under which protocol, and with what uncertainty it is the same.

Even where measurements are physically well-defined, their normative use may remain contested. An index of ecosystem productivity can serve a scientific purpose while failing to represent cultural relations, vulnerability, or independent ecological values. A measure of computational output can facilitate engineering comparison while concealing labor relations or dependencies on infrastructure. The appropriate inference depends on the question and the evaluative commitments governing its application. This distinction will matter for the later treatment of authority over comparative categories; at this stage it establishes that the object of comparison includes the declared function of a measure.

2.5 Histories and the Dependence of Comparative Meaning

Two descriptions may become observationally indistinguishable at a selected resolution while retaining different histories of formation. The histories may affect their responses to intervention, their capacity to continue operating, and the meaning of a common label. A machine-learning model and a rule-based program may give the same answers to a bounded test set while exhibiting different generalization patterns because the procedures generating their behavior differ. Two ecological arrangements may display similar aggregated indicators while depending on different regeneration paths and environmental constraints. These are methodological possibilities, not empirical findings about particular models or ecosystems.

A comparative analysis should therefore specify whether history is relevant to its claim. For a snapshot classification under a complete stipulated task, two systems with matching outputs may legitimately be grouped together. For an explanation of how outputs arise, a history of learning, repair, adaptation, or institutional construction can be essential. For the evaluation of legitimate reliance, it may be important to know who bears the costs of maintaining a capability and whose options are altered by its operation. The adequacy of a comparison is thus indexed to the intended inference, not merely to an observer's preference for rich descriptions.

The inclusion of histories also poses a danger of arbitrary narrative enrichment. An investigator could continually add details until any two systems appeared incomparable. This would undermine the disciplined aim of G4. A relevant historical distinction should therefore have a defensible link to the comparative question: it may change a predicted intervention, a classification under a declared criterion, a material feasibility claim, or an ethical interpretation. Where evidence for such a link is absent, the history can remain recorded as a possible qualification rather than being elevated into an established source of difference.

This approach also avoids equating historical difference with incomparability. Two processes with divergent histories may still satisfy a common conservation relation, support a shared operation, or exhibit comparable causal effects on a carefully controlled variable. Historical dependence identifies a possible condition of interpretation, not a universal prohibition on cross-system inference. The challenge is to establish which historical relations matter, and how strongly, for the particular property being compared.

2.6 Adequacy Relative to a Comparative Question

A comparison is undertaken to answer a question and should be judged against the inferential demands of that question. For G4, this suggests recording the *comparative task* as a tuple of selected targets, interpreted descriptions, regimes, observations, and a proposed inference. A minimal record is expressed in Equation 2.

$$\mathfrak{C}_q = (q; X_a, M_a, X_b, M_b; \kappa_a, \kappa_b; \mathcal{E}_q; \mathcal{J}_q). \quad (2)$$

Here q states the question; $\mathcal{E}_q$ identifies the available evidential procedures and their limits; and $\mathcal{J}_q$ identifies the kind of inference or judgment the investigator seeks to justify. The representation makes

Table 1: Distinct comparative findings and the additional warrants they require.

Finding	Scope of a warranted inference	Additional inquiry
Matching observed outputs	Agreement under a declared observation procedure	Whether unobserved mechanisms or regimes matter
Structural isomorphism	Preservation of relations encoded in the chosen structures	Representational adequacy for the intended targets
Matching reachable outcomes	Feasibility under specified rules	Resource use, joint feasibility, kinetics, and continuation
Shared numerical index	Comparable measured values under an aligned protocol	Construct validity and admissible interpretation
Overlapping interpretation	A common domain of intelligible practice or meaning	Contexts in which translation or coordination fails

no assumption that X_a and X_b share a common ontology or that every component of one description admits a counterpart in the other. In Section 3, the comparative task will be supplemented by partial correspondences and explicit preservation conditions.

Definition 2.2 (Task-relative comparative adequacy). A proposed comparison associated with $\mathfrak{C}_q$ is *adequate relative to the task q* when its selected descriptions and correspondence support the stipulated kind of inference $\mathcal{J}_q$ under declared regimes and available evidence $\mathcal{E}_q$, to a degree appropriate to the inquiry’s stated standards. The standards may be logical for a stipulated structure, predictive and statistical for an empirical model, interpretive for a historical inquiry, or normatively contestable for an institutional application. Adequacy remains open to revision as evidence, regimes, or standards change.

The definition separates the formal completeness of a model from the adequacy of a comparative inference. A fully specified transition system may support exact statements about paths without adequate evidence that it captures the behavior of an empirical target. A partial interpretation may sustain a reliable practical prediction without recovering every relation in the source description. The relevant research question determines the acceptable loss of information. This task-relativity resembles the contextual understanding of scientific representation that attends to representational purposes and interpretation [13]; the present definition organizes those concerns for comparisons among candidate generative systems.

Task-relativity should remain compatible with criticism. An investigator cannot secure an adequate comparison merely by specifying a convenient question whose assumptions ignore foreseeable effects. Where a comparison will guide an intervention, classification, or allocation of standing, its task definition should be open to challenge by those affected. An assessment can be mathematically valid for its chosen inputs while the inputs themselves are incomplete for the proposed decision. Consequently, the record of a comparative task should document both the formal properties of its inference and the reasons for treating its chosen domain as appropriate.

Table 1 organizes several distinctions whose conflation can make a cross-system comparison appear stronger than its warrants permit. The table states possible legitimate findings and the additional question that each finding leaves to be investigated.

2.7 Illustrative Comparisons and the Boundary of the Object

Two schematic comparisons clarify the use of the proposed record. Consider first a pair of stipulated production systems, each with an operation that maps an input marked A to an output marked B . In system a , the input is a reusable symbolic template. In system b , the input is one unit of consumable material. If the question concerns the presence of a one-step transformation from the label A to the label B , the two descriptions agree under an evident label correspondence. If the question concerns the number of outputs jointly available after repeated operation, the systems require different inventory and resource descriptions. The comparative record identifies the step from one-step graph similarity to a stronger production claim as a change in question and in required evidence.

A second comparison concerns a pair of stipulated interpretive practices. Let one practice classify an utterance under a finite vocabulary D_a and another classify a corresponding expression under a partially overlapping vocabulary D_b . A translation may preserve some practical instructions while leaving expressive distinctions without counterparts. A local agreement in response behavior can support a bounded coordination claim. An inference about full semantic equivalence requires additional analysis of contexts, counterfactual uses, and available evidence of interpretation. This resembles the historical debate over incommensurability in the restricted sense that partial failures of translation need not entail complete isolation between frameworks [24, 7]. The example remains hypothetical and supplies no evidence about a particular language community.

The two comparisons also show that the *comparative object* need not be a property intrinsically possessed by one system. It is an organized relation between selected aspects of descriptions, research questions, and evidential conditions. Some properties of the target may be stable across a range of admissible descriptions; others depend on the construction of a specific model or measurement. In either case, the scope of a successful inference should be preserved in the reporting of its result. An inference can be true under one observational protocol and uninformative under another, without logical inconsistency between the two claims.

This section has accordingly supplied the conceptual foundation for the subsequent analysis of partial mappings. A defensible cross-system inquiry begins by declaring its targets and descriptions, its generative regimes, the observables and histories relevant to the question, and the standards of adequate inference. The next step is to determine how selected elements of those descriptions can be related and which relations survive the proposed correspondence. Section 3 develops that step through partial mappings, interpretation, and compositional constraints. The present account retains a provisional status: it specifies a discipline of comparison whose practical utility must eventually be evaluated against established domain-specific research methods and real comparative failures.

3 Relational Correspondence as Partial Mapping

The comparison of heterogeneous generative systems requires a way of relating selected elements of distinct descriptions while retaining the differences that limit an inference. This section develops a preliminary account of *partial relational correspondence*. Its purpose is to make the domain of a proposed interpretation explicit, distinguish different strengths of relational preservation, and establish the conditions under which local correspondences can be composed. The argument proceeds from interpretive selection to typed partial mappings, preservation and reflection of generative relations, quotient constructions, compositional constraints, and the evidential standing of reversible or ambiguous translations. The mathematics uses ordinary set-theoretic and relational constructions. Its intended contribution is the specification of comparative warrants within the broader research programme, rather than a new theorem about mappings.

3.1 The Selection and Interpretation of Corresponding Domains

A comparison commonly begins with an asymmetry of available descriptions. One system may have a mature mathematical model and another may be accessible only through measurements, textual records, or an interpretive exchange. A proposed correspondence therefore relates *selected* portions of each description. The selection should identify both what is translated and what is omitted: configurations, relations, admissibility conditions, measurement procedures, and historical information may enter the comparison at different levels of resolution. The resulting map is relative to a question, rather than an assertion of similarity between whole systems.

Let M_a and M_b be candidate descriptions of the kind specified in Definition 2.1, with configuration sets C_a and C_b . A preliminary domain $A \subseteq C_a$ contains the configurations for which interpretation into M_b is attempted. Its image is a subset of C_b , and the selection may depend on a question q , a regime κ , and the relevant observation procedure. Neither A nor its image needs to exhaust the configurations

of its respective system. The boundary of A matters when a generative operation leaves the selected domain, when a resource needed for that operation has been suppressed, or when an apparent successor is admitted only in a different interpretive context.

A local correspondence may preserve every selected one-step tuple while leaving multi-step paths outside its domain; full-system reachability requires a further closure or path-lifting argument. This boundary is central when the mapping is used to compare prospective abilities.

Selection can also depend on descriptive *types*. A state variable representing a molecular inventory, for example, carries units, admissible values, and a physical interpretation that differ from those of an abstract computational token. A correspondence between inventory and token counts can still be useful for a bounded enumeration question. To support an inference about sustainable production, it would have to retain the material conditions under which counts change. For this reason, a mapping record should include an interpretation of the mapped variables and relational signatures, along with a statement of the questions for which the interpretation has been justified.

There are two directions in which such an interpretation can fail. A selected source configuration may lack a suitable target description; conversely, a target description may contain several possible counterparts whose differences affect the inference. The first motivates a *partial* mapping. The second motivates relations of possible counterparts rather than a uniquely determined function. Treating these as different situations helps preserve the difference between a known absence of correspondence and an unresolved multiplicity of candidates.

3.2 Typed Partial Mappings and Regime Correspondence

For the restricted case in which every selected source configuration has exactly one assigned counterpart, the candidate correspondence is a partial function. Equation 3 gives its elementary representation.

$$f : A \longrightarrow C_b, \quad A \subseteq C_a. \quad (3)$$

Definition 3.1 (Task-indexed partial correspondence). A *task-indexed partial correspondence* from M_a to M_b consists of a comparative question q , a selected domain $A \subseteq C_a$, a function $f : A \rightarrow C_b$, an interpretation record τ for the descriptive types and selected relations, and a partial regime correspondence $\rho : K_a \rightarrow K_b$. The record declares which components of $\mathcal{O}_a, \mathcal{O}_b$ and $\mathcal{H}_a, \mathcal{H}_b$ are pertinent to assessing the comparison, and which claims the proposed correspondence is intended to support. The mapping is *admissible for the task* only to the extent that these declarations and their evidential conditions are satisfied.

The record in Definition 3.1 is deliberately richer than a bare function. A mathematical map between two sets can be well defined while the intended physical or semantic interpretation remains unsupported. For instance, assigning the same integer label to two observations says little about whether the observations are produced by equivalent instruments or refer to corresponding capacities. The interpretation record τ is the place to declare relevant measurement units, permissible renaming of labels, and changes of scale. The regime map ρ similarly requires an explicit account of whether a source regime has an admissible target counterpart. A pair of regimes can be declared corresponding for one generative operation while remaining incomparable for another.

The construction is *typed* in the methodological sense that one records the kinds of elements and operations being related. It does not require every candidate description to be represented in a particular programming-language type theory or categorical formalism. Established work in model theory supplies formal resources for interpreting structures with different signatures, and category theory supplies an account of structure-preserving maps and their composition [16, 20]. Such resources are used here selectively. Whether a scientific or institutional description satisfies the relevant structural axioms remains an additional question.

A total map is the special case $A = C_a$. An injective map distinguishes selected source configurations in its image; a many-to-one map identifies some of them; a bijection $f : A \rightarrow B$ for $B = f[A]$ admits a set-theoretic inverse on the selected domains. These properties have different meanings. Injectivity

can matter to reconstructing distinctions; surjectivity onto a declared image is automatic by construction; and set-theoretic invertibility says nothing, by itself, about the preservation of operational or physical relations. The last distinction becomes important when a comparison is described as *reversible*.

3.3 Preservation and Reflection of Generative Relations

A useful partial mapping may associate configurations successfully while changing the transformations available among them. Generative correspondence therefore calls for a separate declaration of the relations the comparison is meant to preserve. Following the notation of Section 2, let $G_{a,\kappa}^{(m)} \subseteq C_a^m \times C_a$ and $G_{b,\lambda}^{(m)} \subseteq C_b^m \times C_b$ represent admissible $(m + 1)$ -place generative relations. The tuple $(x_1, \dots, x_m; y)$ records a jointly specified set of premises and an admissible outcome under the selected regime. For simplicity, the present discussion assumes that the interpretation record aligns the arity and meaning of these selected operations. An interpretation that changes arity or introduces hidden auxiliary states requires an explicitly richer construction.

Let $\lambda = \rho(\kappa)$ be defined. Equation 4 states *forward preservation* on the selected domain.

$$\begin{aligned} (x_1, \dots, x_m; y) &\in G_{a,\kappa}^{(m)}, \\ x_1, \dots, x_m, y \in A &\implies (f(x_1), \dots, f(x_m); f(y)) \in G_{b,\lambda}^{(m)}. \end{aligned} \tag{4}$$

Definition 3.2 (Local preservation and reflection). A task-indexed partial correspondence is *forward-preserving* for a selected generative relation when implication 4 holds for all tuples in its declared source domain and regime. It is *reflecting* on that domain when the reverse implication holds for tuples whose components have chosen source representatives in A . It is *locally exact* for that relation when both implications hold. These terms describe the corresponding *restricted relations*; they make no assertion about operations involving configurations outside the declared domain.

Forward preservation licenses a limited conclusion: a transformation represented in the selected source relation has a corresponding transformation in the selected target relation. Reflection makes a different assertion: an admitted target transformation between mapped configurations corresponds to a transformation among the specified source configurations. A forward-preserving map can be useful for translating feasible actions or generating candidate hypotheses even when reflection is unavailable. Reflection is needed when the proposed inference runs from the target relation back to the source and would otherwise assign the source capacities observed only in the target.

The restriction to tuples with *all* components in A is consequential. If a source operation consumes an unrepresented resource, a mapping of the visible input and output can preserve a coarse transition while leaving the full generative prerequisites unexamined. Likewise, an operation available in the target under λ may have no source counterpart under κ if the target has additional infrastructure. In both cases, local relational preservation must be distinguished from the stronger claim of equivalent resources, execution costs, probabilities, or continuation conditions. Those stronger claims require additional modeled relations and evidence.

To see the distinction between preservation and reflection, take selected source states $A = \{p, q\}$ and target states $B = \{u, v\}$, with $f(p) = u$ and $f(q) = v$. Suppose the source permits $p \rightarrow q$, while the target permits both $u \rightarrow v$ and $v \rightarrow u$. The map preserves the selected source transition. It fails to reflect the target transition $v \rightarrow u$, since $q \rightarrow p$ is absent from the source relation. A report of “matched behavior” would conceal which direction of inference has been justified. Behavioral-equivalence research makes similar distinctions between one-way simulation and appropriately matched behavior, although the present general-purpose relational record should not be equated automatically with a process-theoretic bisimulation [28].

3.4 Exact Local Correspondence and Quotient Ambiguity

An injective comparison can preserve distinctions among selected configurations. A many-to-one comparison suppresses them, and the suppression may be either harmless for a particular question or decisive

for a generative claim. The following elementary proposition records an exactness condition for the injective case; it is a standard relational-structure observation stated here to delimit the inferential scope of G4's proposal.

Proposition 3.3 (Exactness on selected domains). *Let $f : A \rightarrow B$ be a bijection between declared subsets of two configuration sets, and let $R_a \subseteq A^{m+1}$ and $R_b \subseteq B^{m+1}$ be the induced selected relations. If f both preserves and reflects relational membership, then f is an isomorphism between the two induced relational structures. Conversely, every isomorphism of these induced structures preserves and reflects relational membership.*

Proof. Preservation and reflection together assert that a tuple belongs to R_a precisely when its componentwise image under f belongs to R_b . Bijectivity ensures that every tuple in B^{m+1} has a unique preimage in A^{m+1} . Thus f and its inverse preserve the selected relations, which is the required relational isomorphism. The converse follows from the definition of an isomorphism. $\square$

An exact isomorphism on A and B can coexist with major differences outside those subsets. It also need not preserve an observable that was excluded from the relational signature. Hence the proposition supports a statement about the *induced structures* and the research question that selected them. Additional empirical claims require an interpretation of the structures and suitable evidence that their selected relations characterize the targeted phenomena.

For a many-to-one map $f : A \rightarrow B = f[A]$, a more demanding condition is needed to obtain a faithful *quotient relation*: all tuples with the same componentwise image must agree on whether the selected generative relation holds. Write f^{m+1} for componentwise application and $R_A = G_{a,\kappa}^{(m)} \cap A^{m+1}$. The criterion is stated in Equation 5.

$$R_A = (f^{m+1})^{-1}(f^{m+1}[R_A]), \quad (5)$$

where the inverse image is taken within A^{m+1} .

Proposition 3.4 (Criterion for an exact quotient relation). *For a surjection $f : A \rightarrow B$ and a relation $R_A \subseteq A^{m+1}$, there exists a relation $R_B \subseteq B^{m+1}$ satisfying $R_A = (f^{m+1})^{-1}[R_B]$ if and only if Equation 5 holds. In that case $R_B = f^{m+1}[R_A]$ is uniquely determined.*

Proof. If R_A is the inverse image of R_B , membership is constant on fibres of f^{m+1} , establishing Equation 5. Conversely, if the equation holds, choosing $R_B = f^{m+1}[R_A]$ reproduces R_A by inverse image. Surjectivity of f^{m+1} makes the chosen R_B unique. $\square$

This criterion is stringent, and that is its analytical value. Two configurations can have equal observable labels while differing in the joint transformations available with other configurations. A quotient that replaces the distinction by one label may introduce a transformation supported by only one representative and then attribute it indiscriminately to the other. A weaker quotient can remain useful as an *existential summary*: a target tuple is included when at least one source tuple realizes it. Such a summary should carry an explicit existential interpretation. Equation 5 identifies the stronger circumstances in which the summary can be read as an exact relational correspondence across all selected representatives.

3.5 Composition across Intermediary Descriptions

Comparative research frequently proceeds through intermediary languages or descriptions. A biological system may be related to a dynamical model, that model to a computational representation, and the latter to an empirical index. Each step can be locally defensible while the combined translation is restricted to a smaller domain. Composition must therefore preserve both the mathematical mapping and the evidential conditions under which each component is interpretable.

Suppose $f : A \rightarrow C_b$ is a partial correspondence from M_a to M_b , and $g : B' \rightarrow C_c$ is one from M_b to M_c . The composite is defined only on source configurations whose image under f lies in the domain B' of g . Equation 6 states the resulting restriction.

$$A_{g \circ f} = \{x \in A : f(x) \in B'\}, \quad (g \circ f)(x) = g(f(x)). \quad (6)$$

Proposition 3.5 (Composition of forward-preserving mappings). *Fix an arity m and compatible regimes κ , λ , and μ . Suppose f is forward-preserving from $G_{a,\kappa}^{(m)}$ to $G_{b,\lambda}^{(m)}$ on A , and g is forward-preserving from $G_{b,\lambda}^{(m)}$ to $G_{c,\mu}^{(m)}$ on B' . Then $g \circ f$ is forward-preserving on $A_{g \circ f}$ for the same selected arity and relations. The claim concerns tuples entirely contained in $A_{g \circ f}$ and requires that the regime correspondence assigns κ to λ and λ to μ .*

Proof. For a generative tuple in $G_{a,\kappa}^{(m)}$ with every component in $A_{g \circ f}$, forward preservation under f yields a tuple in $G_{b,\lambda}^{(m)}$ whose components all belong to B' . Forward preservation under g then yields the corresponding tuple in $G_{c,\mu}^{(m)}$. By Equation 6, the resulting components are precisely the images under $g \circ f$. $\square$

Composition is mathematically straightforward under the stated assumptions. The larger epistemic issue is whether intermediate descriptions preserve the distinctions relevant to the final question. Suppose the first mapping reduces two resource inventories to the same visible output count. The second mapping interprets that count as a reusable production capacity. Both steps may support a narrow description of observed outputs, yet the composed interpretation exceeds their combined warrant when applied to repeated physical production. The formal composition theorem offers no justification for such a change of meaning. The interpretation records τ , regime correspondences, and relevant provenance must be checked at the composition interface.

Category-theoretic presentations formalize composition and structure preservation in general settings, and applied category theory develops domain-specific examples involving data, resources, and compositional systems [20, 11]. G4 draws on that mathematical discipline while treating a scientific mapping as an interpreted, empirically or historically situated proposal. When source and target signatures differ or translations introduce intermediate auxiliary states, a correspondence may be better represented by a span, relation, or other structured interface; the simple partial-function case supplies a baseline for specifying the additional requirements.

3.6 Joint Prerequisites and Relationally Coherent Selection

Generative relations involving multiple inputs expose an important limit of elementwise translation. A match between individual input labels is insufficient to establish that the matched inputs can jointly participate in a target operation. Their compatibility can depend on resource budgets, spatial constraints, timing, institutional permissions, or shared context. A mapping that preserves separate unary capabilities may therefore fail to preserve an operation involving their joint availability.

Consider two hypothetical configurations a_1 and a_2 that each support a transformation into an output labelled z under separate regimes. A target description may represent the two operations using an apparently common rule. If the intended comparison concerns a joint operation requiring a_1 and a_2 simultaneously, the evidence for the separate operations leaves the joint premise unresolved. In the notation of Section 2, separate membership in unary relations is a different claim from membership in $G_{a,\kappa}^{(2)}$. The arity, co-admissibility, and regime of the proposed common transformation must remain visible.

A second ambiguity concerns the independence of chosen representatives. Let f identify two source configurations x and x' by the same target configuration u , and suppose only one of them can jointly combine with a third source configuration y . A target relation containing $(u, f(y); v)$ records at most an existentially supported joint possibility unless the fibre-saturation criterion of Proposition 3.4 is satisfied. Substituting x' for x solely because both map to u may produce a false source-level inference. The failure is especially relevant where apparent equivalence is established through output similarity and the research question concerns the conditions of production.

The same logic applies to interpretive and normative examples with suitable changes in the meaning of a relation. Two expressions can each be translated into a shared vocabulary without supporting the inference that they can be substituted within every relational context. Two institutions can recognize ostensibly analogous roles while attaching different powers, obligations, or mechanisms of review. Their

local comparison can be useful, but it is strongest when the relevant combinations and contexts are included in the comparative record. Section 5 will investigate generative compatibility and joint feasibility in greater detail; the present section establishes the mapping-level reason those questions arise.

3.7 Many-Valued Correspondence and Reversibility

A unique assignment is often too strong for an interpretive comparison. A source expression may admit several target readings, and the evidence may underdetermine which is appropriate. The natural elementary representation is then a relation $P \subseteq C_a \times C_b$. For each source configuration x , the set of candidates $P[x]$ records the target configurations consistent with the declared interpretation. The relation can be restricted to a supported subdomain and refined as new evidence becomes available. Its use need not imply that all recorded candidates are equally plausible; quantitative weighting would require another layer of assumptions and measurement.

A relational correspondence requires care when moving from individual pairs to tuples. A list of separate candidate counterparts for $x_1, \dots, x_m, y$ may be drawn from incompatible interpretations or regimes. A purported matched transformation should therefore admit a *jointly coherent witness*: one target tuple that realizes all selected counterparts under the same declared relation and regime. If different observations are each explained using different contextual assignments, their conjunction does not establish one shared translation. This is a methodological requirement to be checked in an application, with its formal implementation depending on how interpretation contexts are represented.

Reversibility can also be stated at several strengths. A bijective function has a set-theoretic inverse on its image. A locally exact relational isomorphism has an inverse preserving the selected relational structure, as Proposition 3.3 records. A translation can be practically reversible for a chosen question when relevant distinctions survive an interpretation and return procedure. Physical or semantic reversibility is stronger still and requires separate evidence about the target phenomena, their states, and the operations available within each setting. A many-valued relation can support reciprocal interpretation even where neither side has a uniquely determined image.

The distinctions are important for comparisons among heterogeneous subjects or practices. A participant's acceptance of a shared term may establish adequate coordination within a specific exchange while leaving further conceptual or normative differences unresolved. A scientific model may be invertible on an observed subspace while losing historical information needed for explaining a system's formation. Reversibility should be attributed to the precise construction and question for which it is warranted. The value of a partial correspondence is its ability to support some conclusions while exposing exactly where others remain unlicensed.

3.8 Comparative Warrant and Documentation of Mapping Limits

The mathematical constructions of this section can organize a comparative inquiry, but they cannot establish on their own that two empirical systems are adequately represented. Their evidential use requires a record that can be challenged, reproduced, and revised. In particular, a study should state how the mapping domain was selected, how types and regimes were interpreted, whether preservation or reflection is required by the intended inference, and whether alternative correspondences would change the conclusion. Table 2 summarizes the principal questions generated by the analysis.

A mapping may therefore support several different conclusions depending on the strength of its verified properties. A forward-preserving assignment can support an inference from a selected source transformation to a target possibility. A locally exact assignment can support a two-way inference for the specified induced relation. An existential quotient can support a statement of possible realization while leaving representative-specific claims unresolved. A coherent composition can support a chain of interpretations when the intermediate domains and regimes match. These are distinct comparative warrants, each of which can be used responsibly when its assumptions and omitted aspects are made explicit.

The philosophical standing of these results remains restricted. Proposition 3.3, Proposition 3.4, and Proposition 3.5 are elementary conditional statements within established relational mathematics. They

Table 2: Conditions to document when evaluating a proposed partial relational correspondence.

Comparative component	Question for verification or revision
Selected domain	Which source configurations and target counterparts are represented, and which are omitted?
Types and interpretation	Which labels, units, operations, and meanings are identified, translated, or left open?
Regime alignment	Under which source and target conditions is the relation claimed to hold?
Direction of inference	Is forward preservation enough, or does the research question also require reflection?
Joint feasibility	Does the evidence support coherent combinations of premises and outcomes, rather than isolated matches?
Information loss	Are distinctions collapsed by a many-to-one map relevant to the intended inference?
Compositional chain	Does every intermediate mapping remain defined and meaningful on the relevant domain?
Empirical warrant	Which observations, interventions, records, or countermodels could support or overturn the proposed correspondence?

clarify the inferential burdens of a comparison without establishing a common ontology among its targets. The empirical utility of this organization remains to be tested against established methods in particular disciplines. A favorable test would show that making domains and preservation directions explicit helps detect a translation failure, formulate an informative competing model, or improve the warranted interpretation of heterogeneous evidence.

The subsequent analysis advances from the existence and scope of a partial correspondence to the properties it preserves. Section 4 will examine invariance under selected transformations, including the difference between exact mathematical invariance, robust empirical regularity, and properties introduced or erased by representational choice. The present section establishes that an invariant claim must name its mapping, its domain, its relevant relations, and the question for which preservation is meaningful. This provides a revisable interface for comparative research while keeping differences among generative histories, material conditions, and interpretive practices in view.

4 Structural Preservation under Selected Transformations

The possibility of a partial correspondence does not by itself determine which aspects of a generative system survive translation. The present section develops a task-relative account of *structural preservation* for selected configurations, relations, transformations, and observations. Its aim is to distinguish exact mathematical invariance, equivariance, regime-dependent preservation, and empirical robustness. The argument proceeds from a declaration of the comparison’s relevant properties to commuting conditions, group actions, conservation, lossy observations, and tolerated measurements. The mathematical statements are conditional applications of familiar results in relational algebra and elementary dynamics. Their contribution to the present inquiry is to specify what can justifiably be inferred across heterogeneous descriptions, and where further domain knowledge is indispensable.

4.1 Preservation Relative to a Declared Comparative Task

The word *invariant* can describe several quite different achievements. A count may remain unchanged when components are relabelled; a conservation quantity may remain constant during an admitted physical transformation; a linguistic relation may survive a stipulated translation; and a statistical coefficient may remain sufficiently stable across sampled groups. Each assertion names a property, a family of transformations, and a range of admissible contexts. Without those specifications, the expression *cross-system invariant* can conceal an unsupported shift from one sort of preservation to another.

Let M_a and M_b be candidate-system descriptions of the form introduced in Section 2. We select domains $U_a \subseteq C_a$ and $U_b \subseteq C_b$, a partial correspondence $f : U_a \rightarrow U_b$ of the kind examined in Section 3, and aligned conditioning regimes (κ_a, κ_b) . The comparative task also selects a signature Σ_*

consisting of the relation and operation symbols needed for the proposed inference. A relation absent from Σ_* is outside the formal preservation claim even if it matters greatly to another investigation. In particular, a mapping that respects a behavioral outcome label need not respect the historical or material conditions through which the outcome was generated.

For a scalar or structured quantity, it is often necessary to translate its value space as well as its configuration. Let $q_a : U_a \rightarrow Y_a$ and $q_b : U_b \rightarrow Y_b$ be quantities with a declared interpretation, and let $\phi : Y_a \rightarrow Y_b$ specify an admissible conversion of units, coding schemes, or categories. An exact value-transport claim is expressed by the commuting condition in Equation 7.

$$q_b \circ f = \phi \circ q_a \quad \text{on } U_a. \quad (7)$$

This identity means that measuring after translation yields the same value as translating an appropriately measured source value. Its empirical use requires an independently defensible interpretation of q_a , q_b , and ϕ . If ϕ merges distinguishable source values, Equation 7 supports equality of translated values while potentially concealing variation in the source. A useful convention is therefore to report explicitly whether the value conversion is injective on the values actually encountered.

Definition 4.1 (Task-relative preserved property). A property selected by a comparative task is *exactly preserved along f* when its truth or value agrees across the declared source domain and target image under an admissible interpretation of its value space. For scalar quantities, this agreement is Equation 7. For a selected relation R of arity m , forward preservation means that every source tuple in R whose entries lie in U_a has an image tuple satisfying the corresponding target relation. A *reflected* property additionally supports inference from the target tuple to the represented source tuple. The claim is restricted to the specified signature, domains, regimes, and question.

This definition deliberately incorporates the distinction between preservation and reflection established in Section 3. Reflection becomes essential when a comparison is intended to justify conclusions about the source from observations of the target. For example, a coding scheme that consistently groups distinct biological reaction states may preserve whether a broad outcome category is attained, while failing to distinguish which resources or intermediate states make that outcome possible. Such a scheme may remain useful for surveillance or descriptive reporting, but its adequacy for mechanistic explanation requires further investigation. The same distinction appears in the philosophy of scientific representation, where structural resemblance and the success of representation for a specified purpose receive distinct treatment [13].

4.2 Commuting Transformations and Dynamical Correspondence

Generative configurations can also be compared through transformations operating upon them. For a simple deterministic benchmark, let $F_a : U_a \rightarrow U_a$ and $F_b : U_b \rightarrow U_b$ be specified admissible updates. We restrict attention to invariant selected domains so that repeated applications remain defined. The assertion that f transports the update is the commutation equation stated in Equation 8.

$$f \circ F_a = F_b \circ f \quad \text{on } U_a. \quad (8)$$

The two paths through the comparison diagram, transforming and then translating or translating and then transforming, consequently agree. The condition describes an *intertwining* or semiconjugacy-type relation in the selected setting. It is stronger than similarity of isolated outputs, and it is weaker than a full dynamical equivalence when f is many-to-one. The latter limitation matters whenever different source states share an observable label while retaining different future possibilities.

Proposition 4.2 (Finite-iteration preservation). Suppose F_a and F_b are self-maps of U_a and U_b , respectively, and $f : U_a \rightarrow U_b$ satisfies Equation 8. Then, for every $n \in \mathbb{N}$ and $x \in U_a$, the finite iterates obey Equation 9.

$$f(F_a^n(x)) = F_b^n(f(x)). \quad (9)$$

Proof. At $n = 0$ the statement is the identity on U_a . If the assertion holds for n , applying Equation 8 to $F_a^n(x)$ yields the assertion for $n + 1$. Induction completes the argument. $\square$

Proposition 4.2 is elementary but provides a useful boundary on interpretation. Finite-iteration preservation follows only after closure of the domains and exact commutation have been assumed. Observational agreement at one instant supplies neither condition. Likewise, an intervention that changes F_a while leaving the chosen target update unaltered can invalidate the equation. For empirical models, one must ask whether measurements support a common transformation rule throughout the range in which the iteration is used. Mathematical existence of an intertwining map cannot substitute for that evidential assessment.

For relational rather than functional generation, the corresponding requirement is preservation of each selected transition or multi-input generative tuple. Finite path preservation then follows along sequences whose intermediates remain in the selected domain, as discussed for relational correspondences in Section 3. The reverse inference needs suitable reflection or path-lifting assumptions. Multi-input generation has an additional feature: configurations representing distinct prerequisites must be jointly admissible in the same regime. An elementwise mapping can preserve observed outputs and still fail to retain the feasibility of combining the inputs. The further analysis of such compatibility belongs to Section 5, while the present distinction concerns the strength of an invariance assertion.

4.3 Symmetry Actions and Cross-System Equivariance

Symmetry provides a disciplined case of invariance relative to transformations. Suppose a group H_a acts on U_a , a group H_b acts on U_b , and a homomorphism $\psi : H_a \rightarrow H_b$ represents the intended relationship between transformation families. A map $f : U_a \rightarrow U_b$ is equivariant with respect to these choices when Equation 10 holds.

$$f(h \cdot x) = \psi(h) \cdot f(x) \quad (h \in H_a, x \in U_a). \quad (10)$$

The equation permits a source transformation to have a corresponding target action. It does not require the two groups to be identical, or their actions to exhaust the transformations important to each system. Equivariance can therefore be useful in heterogeneous comparison while carrying less information than a group-action isomorphism. Its interpretation depends on whether the selected actions represent physical symmetries, descriptive relabellings, or transformations generated within a model.

An observable q_b constant on the relevant H_b -orbits induces an H_a -invariant observable $q_b \circ f$ on U_a whenever Equation 10 is satisfied. Indeed, the target invariance and equivariance give $q_b(f(h \cdot x)) = q_b(\psi(h) \cdot f(x)) = q_b(f(x))$. This familiar argument is valid for the declared actions. Its reverse is generally unavailable when f or ψ omits target states or symmetries. Moreover, invariance of an observable under relabelling should be distinguished from symmetry of the governing transition law. The relationship between physical symmetry, gauge redundancy, and scientific interpretation has been treated extensively in the philosophy of physics [6]; the present analysis borrows the distinction without assuming that an arbitrary GR correspondence is a physical gauge transformation.

An elementary counterexample illustrates the point. Let $x = (x_1, x_2)$ denote the binary occupancy of two independently converting sites, and set $q(x) = x_1 + x_2$. Interchanging the two site labels preserves q . Suppose that a vacant first site becomes occupied at rate 2 and a vacant second site at rate 3, in common arbitrary time units. The rates for the transitions $(0, 0) \rightarrow (1, 0)$ and $(0, 0) \rightarrow (0, 1)$ differ, so swapping sites does not preserve the labelled stochastic generator. The count is permutation-invariant, but the kinetics lacks that permutation symmetry under these conditions. One can retain the count as a descriptive common quantity without inferring dynamically equivalent sites or an exactly Markovian count process.

This example also explains why the nature of the selected transformation matters. A similarity under a representational relabelling may be exact for a chosen quantity and irrelevant to a scientific question about rates, causal mechanisms, or resource use. An empirical claim about a physical symmetry needs independently justified dynamical structure and observations. A comparative claim about equivalent institutional or linguistic roles requires analogous care about meanings, authorizations, and interpretive

conventions. The useful common lesson concerns evidential boundaries, rather than an identity between these different domains of inquiry.

4.4 Conserved Quantities and Information Lost in Translation

Conservation provides a second, more restrictive notion of invariance. For a transition relation $T \subseteq U \times U$, a quantity $q : U \rightarrow Y$ is conserved along T when $q(x) = q(x')$ for every $(x, x') \in T$. This definition refers to the selected transitions; it does not identify their physical source or assert conservation under unexamined transformations. In a physical application, the relevance of q must be justified through the governing theory and its interpretation. In other domains, an invariant might concern permitted balances, preserved permissions, or formally maintained properties, with the meaning of the claim correspondingly altered.

Suppose a source transition T_a is forward-preserved by f into a target transition T_b . Let $q_a : U_a \rightarrow Y_a$ and $q_b : U_b \rightarrow Y_b$ satisfy the value-transport equation. Target conservation then yields a limited source result. Its precise strength depends on whether the conversion ϕ retains distinctions among values.

Proposition 4.3 (Pullback of a conserved quantity). *Assume that $f : U_a \rightarrow U_b$ maps each transition in $T_a \cap (U_a \times U_a)$ to a transition in T_b . Let q_a, q_b , and ϕ satisfy Equation 7, and suppose q_b is conserved along T_b . Then $\phi \circ q_a$ is conserved along the selected source transitions. If ϕ is injective on $q_a[U_a]$, q_a itself is conserved along those transitions.*

Proof. For a selected transition $(x, x') \in T_a$, forward preservation gives $(f(x), f(x')) \in T_b$. Conservation of q_b and Equation 7 imply $\phi(q_a(x)) = \phi(q_a(x'))$. Injectivity of ϕ on the image of q_a gives $q_a(x) = q_a(x')$. $\square$

Proposition 4.3 identifies an ordinary source of apparent invariance. Let $q_a(x) = 0$ and $q_a(x') = 1$ for a source transition $x \rightarrow x'$, and let ϕ send both values to the same broad category. The translated quantity is constant although q_a changes. A target relation containing the image transition can therefore conserve the recorded category without preserving the finer source distinction. A coarse classification may remain appropriate for a task that asks only whether the broad category persists. It is inadequate for a task that claims exact conservation of the original quantity.

The reverse direction has a complementary limitation. If q_a is conserved along selected source transitions and f covers only part of the target's transition system, then target transitions outside that image may change q_b . Conservation within a represented submodel therefore warrants conservation only within that submodel. This point applies even when the local correspondence is mathematically exact. Evidence about domain coverage, external influences, and regime changes is needed to expand the claim.

4.5 Representational Quotients and New Apparent Invariants

A quotient can generate a convenient, stable descriptor while erasing distinctions relevant to mechanistic comparison. Let $f : U_a \rightarrow U_b$ identify several source configurations, and suppose R_b is formed existentially from the images of selected source tuples. Even when each source relation is mapped forward, the target may admit additional apparent substitutions because several incompatible representatives share a label. Section 3 formalized exactness through fibre saturation; in the present context the question is whether a proposed invariant survives representative choice and available operations.

Two requirements should be separated. First, a candidate quantity q_a must be constant on each fibre of f if it is to descend to a well-defined target quantity $\bar{q}$ satisfying $q_a = \bar{q} \circ f$. Second, the desired generative inference may require the relation itself to descend exactly, which is the stronger relational condition given by Proposition 3.4. Fibre-constant quantities alone say nothing about whether separately admissible operations can be combined or whether the quotient has introduced paths unsupported by any coherent source history.

For example, imagine an observational label that records whether a production system can yield at least one acceptable output. Two underlying configurations may receive that same label although one relies on an expiring input and the other on a continuously available resource. The label may remain

stable across repeated reporting periods; its stability does not demonstrate invariance of the capacities required to reproduce the output. A second report that records only the presence of a shared institutional title can similarly conceal different rules governing access, challenge, and authority. Such illustrations should be treated as methodological thought experiments rather than empirical claims about particular organizations or populations.

The direction of abstraction also matters. Introducing a more detailed target description can split a formerly invariant class, while a coarser description can create an invariant category by suppressing relevant differences. Neither outcome determines which description is more adequate without a research question and a specified evidential target. The task-relative perspective therefore treats the visibility of an invariant as a joint result of the phenomenon, the representation, and the declared transformations. It also allows a failed invariance test to be informative: the failure can identify a condition, history, or variable that an earlier comparison omitted.

4.6 Approximate Preservation and Stability of Comparative Rankings

Exact commuting conditions are demanding in empirical work. Measurement uncertainty, sampling variation, model approximation, and incomplete calibration can yield estimates that are informative within a declared tolerance. Approximate preservation should be stated in terms of an error measure suited to the comparison’s task, rather than as a vague resemblance. For real-valued quantities whose units have been aligned through ϕ , suppose Equation 11 holds for the selected domain.

$$|q_b(f(x)) - \phi(q_a(x))| \leq \varepsilon \quad (x \in U_a). \quad (11)$$

Here $\varepsilon \geq 0$ is a declared upper bound, justified by calibration, formal approximation, or a stated uncertainty model. In a statistical application, the bound may hold only with specified confidence or probability; the deterministic equation would then require the corresponding qualification. A low average discrepancy is also different from a uniform worst-case guarantee. These differences affect what one can infer about particular configurations and should be reported in the comparative protocol.

Proposition 4.4 (Margin-stable comparison of quantities). *Assume Equation 11 and let $x, y \in U_a$. If $\phi(q_a(y)) - \phi(q_a(x)) > 2\varepsilon$, then $q_b(f(y)) > q_b(f(x))$.*

Proof. The error bounds yield $q_b(f(y)) \geq \phi(q_a(y)) - \varepsilon$ and $q_b(f(x)) \leq \phi(q_a(x)) + \varepsilon$. The strict margin assumption gives the asserted strict inequality. $\square$

Proposition 4.4 is a straightforward consequence of the triangle inequality. It specifies an elementary way for a comparative ranking to survive bounded translation error. The condition is sufficient, and it concerns the selected numerical quantity rather than a system’s overall generative status. If ϕ is order-preserving, the original numerical ordering can also be related to the target ordering; if ϕ reverses order or merges values, the interpretation changes. At the threshold 2ε , the bounds allow a tie, and smaller margins may permit reversal. The same reasoning provides limited certificates for threshold classifications when the distance to a decision boundary exceeds the relevant error bound.

Measurement invariance in empirical social and psychological research supplies a more developed methodological comparison. It concerns conditions under which constructs and indicators can be compared across groups or occasions, with distinct levels of measurement equivalence and corresponding inferential limits. The analysis of Reise, Widaman, and Pugh [27] and the methodological review of Vandenberg and Lance [31] illustrate why a shared numerical response scale alone need not license substantive comparisons. The present abstract error bound does not replace tests of construct validity, group-specific mechanisms, or measurement-model adequacy. It merely identifies one formal condition for an explicitly selected scalar inference.

4.7 Comparative Tests of Preservation Claims

A statement of structural preservation should be accompanied by observations or models capable of contesting it. The appropriate evidence differs across physical, computational, historical, and interpretive

Table 3: Selected preservation claims and the additional evidence needed for warranted interpretation.

Claim type	Formal requirement	Empirical or interpretive check
Value transport	$q_b f = \phi q_a$	Units, definitions, calibration, and information lost through ϕ
Relational preservation	Source relation maps into target relation	Domain coverage, regime agreement, and omitted tuples
Dynamical correspondence	$f F_a = F_b f$	Transitions beyond sampled instances; changes following interventions
Equivariance	$f(hx) = \psi(h)f(x)$	Meaning of actions; difference between relabelling and physical symmetry
Conservation	Quantity constant on selected transitions	Physical or institutional law and whether relevant transitions are included
Approximate ranking	Uniform error bound and gap $> 2\varepsilon$	Error model, calibration, comparison margins, and uncertainty coverage

investigations, yet several common questions can be asked. Which configurations are covered? Which transformations are admissible? Are the observed quantities calibrated to comparable meanings and units? Does preservation concern a relation, a trajectory, an output, or a conservation law? Is the asserted property exact, boundedly approximate, or merely suggested by a finite sample? Table 3 organizes these questions as a comparative audit rather than a general criterion of ontological identity.

A useful comparative research design can proceed by declaring the intended inference, identifying at least one candidate mapping, testing the selected preservation condition, and recording countermodels that reproduce the observed similarities while violating the intended conclusion. Such a design can also compare alternative representations and ask whether the same finding survives plausible changes of regime and measurement. The construction becomes particularly valuable when a failed mapping can be associated with a specific missing premise, such as a resource constraint, an unmeasured environmental influence, or a historically produced interpretive convention. The present paper has not yet established through independent studies that this procedure improves comparative practice; that claim remains a research proposal.

A further distinction concerns who selects the comparison. A researcher can identify a mathematically elegant invariant while affected participants regard the omitted differences as central to their practices or obligations. Equally, a community may employ a locally meaningful equivalence that has limited transferability to another scientific or institutional setting. These possibilities call for transparency about the authority, purposes, and consequences attached to comparative categories. Their normative and procedural implications will receive separate examination in Section 10; they should not be derived from the existence of a commuting diagram.

The present section therefore supplies a hierarchy of warranted formal statements without ranking the systems themselves. A partial mapping may preserve selected relations, a commuting transformation may preserve finite iterations, an equivariant comparison may preserve an orbit-invariant observable, and an approximate scalar transport may preserve rankings separated by adequate margins. Each result remains conditional on its own assumptions. The next section turns to generative compatibility across differing conditions, where joint prerequisites, consumable resources, and the simultaneous validity of mapped operations become the principal objects of comparison.

5 Generative Compatibility across Distinct Conditions

A relational correspondence can preserve selected transformations while leaving open whether those transformations can be realised together. The present section investigates this additional question through the joint conditions of generative activity. Its objective is to distinguish individually admissible operations from jointly feasible arrangements, and to identify the material, institutional, and representational

assumptions required to compare their compatibility across heterogeneous systems. The discussion begins with a conceptual account of compatibility relative to declared conditions, introduces an inventory-sensitive model of generation, proves two elementary results about feasibility and composition, and then examines regime changes, historical dependencies, and practical comparison protocols. These formal constructions are deliberately restricted: they clarify particular inferences about resources and operations while leaving the interpretation of a system’s generative significance open.

5.1 Compatibility as a Conditioned Comparative Relation

A partial mapping from one system to another, as developed in Section 3, identifies selected configurations and may preserve a relation describing how an outcome follows from a set of premises. Structural preservation under transformations, examined in Section 4, strengthens a comparison when a map intertwines dynamics or respects a specified quantity. Neither result, taken alone, settles whether operations share an admissible regime, draw upon the same finite resources, or can take place within one coherent configuration. The relevant analytical object is therefore *generative compatibility under declared conditions*. Compatibility is indexed by the operations to be combined, their demands, the available resources, and the type of arrangement under consideration.

One can distinguish at least three tasks that sometimes receive the same description. A *possibility comparison* asks whether an outcome can be obtained under some admissible conditions. A *joint-feasibility comparison* asks whether several outcomes can be obtained within a single resource-constrained realisation. A *continuity comparison* asks whether their production can be sustained under a succession of relevant conditions, including replenishment and wear. These tasks require progressively different information and admit different evidential tests. Observing two outcomes in separate trials may establish possibility for each; it generally supplies insufficient evidence for their joint production under one unchanged initial endowment.

For the present analysis, a *regime* specifies the admissibility and interpretation of operations as well as relevant boundary conditions. An inventory or resource state specifies what remains available within that regime. A comparison between regimes must record whether the same resources, environmental conditions, or institutional permissions are assumed. An operation can be technically specified yet disabled by a shortage, restricted by a binding obligation, or unavailable because a required participant has withdrawn. The term *compatibility* names a relation between selected activities under these conditions, rather than a property of an entire subject detached from its relations.

The distinction is useful across the cases considered in this paper. A computational service may expose two functions while allocating insufficient memory or computing time to run them together. A laboratory protocol may permit each reaction in isolation while giving the reactions a shared, consumable reagent. A collaborative research arrangement may describe several participating institutions as able to contribute, while their mandates, data permissions, and scheduling constraints make simultaneous performance unavailable. The examples express different mechanisms; a common comparative vocabulary does not imply that material scarcity, software scheduling, and institutional permission are instances of one physical law. Their conditions of validity must be established separately.

5.2 Joint Prerequisites and Resource-Sensitive Generation

To examine physical and computational resources within a transparent formal setting, let P be a finite set of resource types and let a *marking* $m \in \mathbb{N}^P$ record the available quantity of each type. A finite collection $\mathcal{R}$ of rules is specified by input and output vectors $\alpha_r, \beta_r \in \mathbb{N}^P$. The present interpretation treats these inputs as consumable or temporarily unavailable quantities and the outputs as quantities subsequently available. A rule fires when its input is covered by the marking. Equation 12 gives the elementary firing relation.

$$m \xrightarrow{r} m - \alpha_r + \beta_r \quad \text{provided} \quad m \geq \alpha_r \text{ componentwise.} \quad (12)$$

Equation 12 is an ordinary marked Petri-net construction, included as a model of resource-sensitive generation rather than an original mathematical innovation. Its established history and compositional

developments offer a relevant comparison to the present project [4]. Resource-sensitive logical distinctions also have a different, proof-theoretic treatment in linear logic [14]. The physical significance of a token, and whether a given resource can genuinely be consumed, restored, duplicated, or shared, depends on the application; the formal descriptions are complementary tools rather than universal substitutes for empirical investigation.

The inequality in Equation 12 distinguishes the availability of a premise from its mere occurrence in a rule description. It also provides a concrete setting in which compatibility is more demanding than individual enablement. Given a finite multiset of intended firings represented by $n \in \mathbb{N}^{\mathcal{R}}$, suppose they must begin together, with every input reserved before any output can be used. Equation 13 characterises their joint input feasibility under this *one-round reservation* convention.

$$\text{Joint}_m(n) \iff m \geq \sum_{r \in \mathcal{R}} n_r \alpha_r. \quad (13)$$

Definition 5.1 (Conditioned generative compatibility). For a specified inventory model, regime κ , initial marking m , and finite intended collection n of rule firings, the collection is *jointly compatible in one round* when its rules are admitted by κ and the componentwise reservation condition in Equation 13 holds, together with any separately declared global restrictions. A collection is *sequentially feasible* when there exists an admissible ordering of its firings, each enabled at the marking resulting from the preceding firings. These are task-indexed concepts. The definition does not equate joint one-round compatibility with sequential reachability, viability over an unbounded horizon, or normative permission.

One-round reservation is intentionally stricter than a sequential schedule that can use intermediate outputs. It is also more restrictive than arrangements where a catalyst or institutional credential remains reusable among operations: such cases call for explicit read-only, inhibitory, capacity, or scheduling conditions in a richer model. The comparative record should specify which kind of use is being claimed. The phrase *same resource* may refer to a single consumable object, a reusable instrument with finite processing capacity, an access right that permits several uses, or a replenishable stock. Conflating these cases can produce apparent incompatibilities as readily as overlooking genuine conflicts.

5.3 Marginal Possibility and Joint Attainability

A basic counterexample shows why the marginal reachability of two outcomes is a weaker statement than their joint attainability. Consider three species A , B , and C , with initial marking $m_0 = (1, 0, 0)$ and two rules $A \rightarrow B$ and $A \rightarrow C$. Both terminal configurations $(0, 1, 0)$ and $(0, 0, 1)$ are reachable from m_0 by a single firing. The configuration $(0, 1, 1)$ would require two output units although the permitted rules each consume the sole initial A unit and supply one product. It is therefore unreachable from this marking. The model has two individually reachable alternatives and one unsatisfied joint target. It does not establish any general empirical claim about real chemical production without identifying the relevant reactions and conservation laws.

The comparison can be expressed in terms of an outcome predicate. Let V_B identify markings containing B and V_C identify markings containing C , with $\text{Reach}(m_0)$ the markings reachable by finite firing sequences. The two marginal claims are $\text{Reach}(m_0) \cap V_B \neq \emptyset$ and $\text{Reach}(m_0) \cap V_C \neq \emptyset$. The joint claim requires $\text{Reach}(m_0) \cap V_B \cap V_C \neq \emptyset$. In the stated example the first two claims hold and the joint claim fails. The distinction is set-theoretically elementary, yet it matters when a comparison uses a list of separately reachable products as a proxy for the system’s ability to maintain them together.

The converse caution is equally important. If two outcomes are found together, that observation does not identify the generative route that produced them; an earlier stock, an external source, or an unobserved transformation may explain their coexistence. Establishing a claimed resource mechanism requires evidence about inventories, flows, initial conditions, and the time or event scale of observation. A snapshot may show the result of a feasible trace without revealing which trace occurred. This point also anticipates Section 2’s distinction between the target system and the observer’s descriptive commitments.

Proposition 5.2 (One-round resource criterion and its limitation). *Suppose the only constraints on a collection of simultaneous firings are the availability of consumable inputs α_r in a single initial marking m , the admissibility of the selected rules, and the requirement that their inputs be reserved before any output becomes available. The collection is feasible precisely when Equation 13 holds. This is a one-round, simultaneous-reservation criterion: it does not characterize arbitrary sequential execution, catalytic reuse, or outputs becoming available between firings. In general, the individual enabling of each selected firing does not imply this joint condition.*

Proof. The stipulated one-round interpretation requires a separate allocation of α_r for each occurrence of rule r ; summing the required inputs gives the componentwise necessary condition. If the inequality holds, the marking supplies those allocations and the declared assumptions permit the joint firing. For the final assertion, let m contain one token of type A and let two distinct firings each require one A token. Each is individually enabled, while their combined demand exceeds m_A . $\square$

Proposition 5.2 concerns a deliberately restricted operational convention. It should not be mistaken for a novel feasibility theorem or extended automatically to concurrent systems that permit resource sharing, fractional quantities, alternative pathways, or intermediate replenishment. The value of the formal result in G4 is diagnostic: it identifies which additional premises a comparison must contain before an inference from separate generative possibilities to joint capability is warranted. In applications, this can be tested by a detailed resource ledger, an executable scheduling model, or an intervention that holds the relevant supply conditions fixed.

5.4 Independent Composition and Shared Constraints

Compatibility can sometimes be established compositionally. The strongest elementary case arises when two subsystems have disjoint resource types, local rules, and independent admissibility conditions. Under these assumptions, each subsystem has its own marked transition system, and the composite has product markings. Changes in one component leave the other unchanged. A route that is feasible in the first component can then be interleaved with a feasible route in the second without resource competition. Equation 14 states the resulting product form of reachability.

$$\text{Reach}_{1 \oplus 2}(m_1, m_2) = \text{Reach}_1(m_1) \times \text{Reach}_2(m_2). \quad (14)$$

Proposition 5.3 (Reachability under independent interleaving). *Let two marked rule systems have disjoint resource sets and disjoint local transition rules, and suppose their composite permits arbitrary finite interleavings of locally enabled transitions with no shared constraints. Then Equation 14 holds for any initial product marking (m_1, m_2) .*

Proof. A finite composite trace consists of firings from the first or second component. Projecting its firings onto either component yields a valid local trace, so the composite endpoint belongs to the product on the right. Conversely, any two finite local traces may be interleaved or concatenated because each firing changes only its own component and the hypothesis excludes shared restrictions. Their endpoint is therefore reachable in the composite. $\square$

Proposition 5.3 is a standard independence property. It isolates the precise assumptions supporting a positive inference from separate local reachability to joint reachability. A single shared stock, a shared deadline, a global power limit, or a synchronization requirement may invalidate the factorisation. The product construction is therefore a useful *null model* of independence against which a proposed cross-system composition can be examined. Its failure can indicate a hidden interface, an omitted constraint, or a misdescription of the relevant configuration space.

The difference between disjoint composition and composition through interfaces deserves special attention. A networked system may export outputs from one subsystem as inputs to another, permit shared access to a capacity, or require matched operations. Such systems can still be modeled compositionally, provided the connecting interface specifies how resources, transitions, and conditions are

identified. Established accounts of open Petri nets make these interface commitments explicit and distinguish operational composition from reachability summaries [4]. Similarly, the compositional vocabulary of category theory can track mappings that respect a selected kind of structure [11]. The particular semantics of the interface must be fixed before a mathematical composition is interpreted as a feasible physical or institutional arrangement.

The present analysis does not assign a single preferred abstraction to every interface. An electrical supply constraint, a collective research agreement, and a shared ecological habitat involve different mechanisms and kinds of evidence. The common methodological requirement is to represent interactions that can defeat independent composition. In some cases, the relevant interface is a physical flow, in others a capacity allocation, a permission, or a coordinated sequence of decisions. Comparisons should preserve that diversity while identifying whatever restricted structure genuinely transfers.

5.5 Regime Translation and the Continuity of Comparative Conditions

Cross-system correspondence becomes fragile when a comparison silently changes the admissibility conditions that make its source operation possible. Let κ_a and κ_b be declared regimes, with the partial regime correspondence ρ from Definition 3.1. An operation admitted in κ_a may have a representational counterpart in κ_b while relying on a different enabling inventory, institutional approval, or environmental supply. Forward preservation of its state transition is then a claim about selected tuples; compatibility across complete activities requires additional information about their enabling conditions and continuation.

A workable description can augment configurations with resources and regime information, so that an operational state has the form $s = (c, m, \kappa)$. A proposed comparison then records a partial map on the admissible operational states and, when pertinent, the resource and regime interpretations that accompany it. If the intended inference concerns an entire finite trace, each successive mapped state must remain in the chosen domain and each mapped operation must remain enabled under the target regime. This is stronger than observing that isolated input and output configurations can be assigned corresponding labels.

A simple induction explains the relevant sufficiency claim. Suppose every selected source operational transition between states in the comparison domain has a corresponding target transition, the initial operational states correspond, and the image of each source successor remains in the domain throughout the trace. Then the mapped trace exists in the target transition system, by applying the one-step preservation property successively. The proof resembles the composition argument in Section 3; its significance here lies in making the *complete operational state*, including enabling conditions, the object mapped. Merely comparing projected outcome labels may conceal an exhausted resource or a changed permission between steps.

The converse inference would require a suitable reflection or lifting condition, which may fail because a target system has additional sources of replenishment, more generous admissibility conditions, or a greater choice of implementations. A simulation of a source process by a target therefore need not establish reciprocal interchangeability. In more demanding applications, a relation between operational states may be preferable to a single-valued mapping, particularly when several target states can instantiate one source description. Such a relation must still declare what counts as matching a transition and which differences matter to the comparative question.

Compatibility across changing conditions also raises a historical issue. A rule might be enabled today by a stock accumulated through an earlier process or by an entitlement established through an institutional history. Its present success may depend on conditions absent from the instantaneous configuration. A state representation containing only current visible outputs can therefore omit the provenance that makes continued operation possible. For that reason, longer-horizon comparisons should register stocks, replenishment, contractual or ecological dependencies, and the assumptions used to treat a process as repeatable. Section 6 will develop historical provenance as a separate dimension of comparative inference.

5.6 Material, Institutional, and Ecological Interfaces

The resource models clarify particular aspects of joint feasibility, but a comparison of heterogeneous subjects or systems involves relations that exceed a token ledger. Material infrastructures, legal permissions, social recognition, and ecological dependencies can each function as conditions enabling a practice. Their inclusion requires a description adequate to the question, with their respective evidential and ethical limits stated. The same formal indicator of availability can correspond to a measurable stock, a temporary authorization, or an interpretive convention; agreement of the indicators establishes neither identity of those mechanisms nor equal consequences for participants.

Consider, as a hypothetical institutional illustration, two organizations that separately advertise access to the same data repository. The organizations may each maintain legitimate credentials, but a shared licensing arrangement may prohibit simultaneous extraction or impose a single rate limit. If the comparison asks whether each organization can execute one query, marginal access may suffice. If the question concerns their ability to conduct a coordinated analysis, the shared capacity, permitted uses, and scheduling rules become part of the generative conditions. A corresponding pair of state labels offers little warrant until these interface conditions have been examined. The example is an analytical construction and does not describe a real institution.

An ecological illustration differs in its physical and normative character. Two populations may each persist under separate resource conditions, while their combined presence in a shared habitat alters nutrient flows, space, or exposure to risk. Determining whether their coexistence is feasible requires a model or empirical inquiry into interaction, replenishment, and relevant timescales. The fact that a mathematical model permits an allocation does not itself answer whether that allocation is sustainable, ethically acceptable, or appropriately governed. In human–AI systems, a similar distinction appears between computational ability and the authority to deploy an operation on data belonging to others. In each case, the formal model identifies questions to investigate; domain-specific science, law, and stakeholder interpretation determine whether its assumptions hold.

Such heterogeneity cautions against defining a universal index of compatibility. A binary operational test may be adequate for a finite scheduling task but too coarse for ecological viability or institutional consent. Where evidence is incomplete, a comparison can return a conditional result or a set of possible outcomes rather than a definitive equivalence claim. In particular, a lack of observed incompatibility should be distinguished from a demonstrated absence of conflict under a sufficiently comprehensive test. This distinction preserves the epistemic role of unresolved cases while helping researchers identify the next measurements, documents, or interventions required.

5.7 Comparative Evaluation of Joint Feasibility

An evaluation of generative compatibility should begin with the intended inference and the scope of the comparison. The researcher can then state a regime, the operational objects being related, the proposed mapping, and the conditions necessary for simultaneous or successive performance. The resulting claim may be exact within a formal model, supported by observation under stated confidence, or a provisional hypothesis requiring further inquiry. Table 4 records several common inferential steps and the observations most likely to challenge them.

Table 4 should be read as a menu of discriminating questions rather than an all-purpose instrument. It invites several concrete research designs. In a formal comparison, one can calculate reachability for isolated and joint objectives, document the enabling markings, and identify a minimal resource conflict. In a physical application, one can measure the relevant inventories and fluxes, test conservation assumptions, and compare model predictions with controlled interventions. In an institutional comparison, one can examine written permissions, resource-allocation practices, and participants' ability to contest a declared mapping of capabilities. These procedures may generate mutually informative evidence while retaining different standards of validity.

A rigorous comparative report should also record countermodels. The one-unit example of Section 5.3 supplies a countermodel to an unrestricted inference from separate possibility to joint feasibility.

Table 4: Levels of generative compatibility and evidence required for cross-system comparison.

Target inference	Declared model requirement	Relevant challenge or observation
Individual possibility	Separate reachable output for each activity	Distinct initial stocks or incompatible regimes across trials
Joint one-round feasibility	Common inventory covers total reserved inputs	Shared consumables, capacity limits, synchronisation
Sequential compatibility	One valid ordered trace with updated markings	Intermediate resource depletion or hidden enabling changes
Independent composition	Disjoint resources and local rules, unrestricted interleaving	Shared interfaces, quotas, permissions, deadlines
Cross-regime process translation	Operational-state mapping preserves enabling and selected steps	Incomplete regime translation, replenishment differences
Continuing coexistence	Declared replenishment, viability and relevant time horizon	Cumulative losses, ecological feedback, revocable rights

The independent-product result of Proposition 5.3 supplies a positive benchmark under explicit independence assumptions. The contrast helps locate the burden of proof: a claim of compatibility needs evidence for the conditions that connect the proposed operations, while a claim of incompatibility needs evidence that the apparent obstacle persists under the permitted alternative paths, replenishment, and institutional arrangements. This balanced requirement is important when comparative findings may influence access, cooperation, or allocation decisions.

The proposed approach remains an evaluative framework rather than an empirically established general theory. The propositions employ standard finite-state and resource-sensitive mathematics; their novelty is not asserted. The philosophical contribution is to identify compatibility as an additional, task-relative warrant within cross-system comparison and to keep that warrant separate from observational similarity, structural preservation, and normative justification. The next section examines the historical provenance of comparable configurations, where different paths of formation and inherited dependencies can alter the significance of otherwise similar present possibilities.

6 Historical Provenance in Comparative Inference

Comparisons between heterogeneous generative systems frequently begin from currently available descriptions, measured outputs, or institutional arrangements. Yet configurations that agree under these descriptions can retain different possibilities because the relations through which they were formed have altered hidden resources, permissions, dispositions, or modes of interpretation. Historical provenance provides a way to assess this possibility. The present section specifies the analytical roles of formation histories, distinguishes evidence of prior events from claims about causal dependence, and asks when a present-state correspondence supports inferences about subsequent generative activity. It develops a finite history-sensitive transition model, proves a restricted criterion for preserving observable continuations, and examines practical cases and evidential limitations. The formal model uses a declared succession of recorded events for analysis; it presupposes neither a universal temporal order nor an ontologically privileged conception of history.

6.1 Formation Histories as Comparative Evidence

A *formation history* is a supported account of events, transformations, inheritances, and constraints relevant to the emergence of a selected configuration. Its content depends on a question: a technical comparison may require manufacturing and calibration records, whereas an institutional comparison may require changes in authorization, responsibility, and access to records. The history is therefore a selected and revisable evidential object. Its admissible components may include measured changes, documentary

records, preserved material traces, witness reports, or inferred but uncertain intermediate configurations. A report about an event and the event's physical occurrence are different objects of inquiry, even where reliable evidence connects them.

The history of a generative configuration can matter in several distinguishable respects. *Material inheritance* concerns resources, wear, stored energy, or retained structures. *Relational inheritance* concerns dependencies on other participants and infrastructures. *Institutional inheritance* concerns enforceable permissions, obligations, and procedures. *Interpretive inheritance* concerns conventions that give current symbols or practices their operative significance. These dimensions may coexist, but their mechanisms and evidence differ. Their relevance must be demonstrated relative to the comparative conclusion under examination.

Path-dependence research supplies established reasons to take such variation seriously. Arthur's increasing-returns model relates adoption histories to the eventual selection of competing technologies, and Pierson explains how increasing returns can make sequences and contingent junctures consequential in political development [3, 25]. These works support a bounded methodological motivation. Their particular mechanisms should be tested within the relevant economic or institutional domain; the present inquiry does not extend increasing returns indiscriminately to biological, linguistic, or artificial systems. In each field, a historical claim requires a specified process through which earlier conditions remain consequential.

Provenance may be represented as a sequence where the admissible events have an agreed order, or as a partially ordered network where only some precedences are warranted. The second representation may be more appropriate for distributed institutional operations, interacting ecological processes, or heterogeneous observational records. Selection of a sequence in the formal illustrations below is a convenience for finite derivations. The philosophical conclusions concern relations of generation and evidence that can also be represented without a single comprehensive clock or chronology.

6.2 Present Correspondence and Historical Indeterminacy

Let a current description be supplied by an observation map o and a configuration space C . Two configurations c and d may satisfy $o(c) = o(d)$ while differing in additional variables that regulate their admissible transformations. Equality of observation can therefore certify a common present appearance while leaving the relevant formation histories unresolved. Even exact equality of selected structural attributes permits historical differences, since several transformations may lead to one observable state. The question of whether those differences matter for future comparison turns on the selected operations and their enabling conditions.

The distinction builds on Sections 3 and 5. A local map can preserve current relational attributes, and separate operations may each be feasible under their own declared resources. A stronger inference concerns whether matching present descriptions license matching sets of subsequent actions *within one coherent history and regime*. This requires attention to what has already been consumed, which rights or capacities remain available, and how the admissibility of a future operation depends on previous events.

Consider two hypothetical measuring devices with the same current display and nominal specification. One has undergone a recent calibration procedure, while the other has an unrecorded exposure that changes its response to a future perturbation. Their present displays can agree, yet predictions about measurement reliability require additional information. The example establishes a logical possibility rather than a factual claim about a class of instruments. Correspondingly, two administrative roles might carry the same title while differing in powers inherited through prior delegations. In both cases, the relevant comparison extends beyond the visible label toward the history-dependent rules governing what the configuration can do.

Historical indeterminacy must also be distinguished from genuine historical dependence. A configuration may have multiple recorded paths of formation with identical present capabilities and identical responses to all interventions relevant to the inquiry. In that situation the paths can remain distinct for archival or interpretive purposes while making no difference to the selected predictive task. An adequate

framework should allow both possibilities rather than infer path dependence from the mere existence of different narratives.

6.3 A History-Sensitive Model of Generative Continuation

A restricted transition model makes the distinction precise. Fix a regime κ , a set of admissible finite histories $\mathcal{H}_\kappa$, an alphabet A of selected actions, and a set Y of observed outcomes. A history $h \in \mathcal{H}_\kappa$ is a record of configurations and events in an admissible local order. Write $h \xrightarrow{a} h'$ when action a has an admissible successor history h' . Let $q : \mathcal{H}_\kappa \rightarrow Z$ record the current comparative description, with codomain Z . The map q may omit earlier operations, hidden stores, permissions, or other provenance variables. For the moment, its target is a formal observational description, rather than a claim about the uniquely correct state of the world.

Definition 6.1 (Task-relative provenance adequacy). For a set of selected actions $A_0 \subseteq A$ and a current-description map q , a description is *one-step continuation adequate* on a supported domain $H_0 \subseteq \mathcal{H}_\kappa$ when, for any $h, k \in H_0$ with $q(h) = q(k)$ and any $a \in A_0$, the sets of observable successor descriptions coincide, provided the relevant successors belong to H_0 . Formally, the requirement is given in Equation 15.

$$\{q(h') : h \xrightarrow{a} h', h' \in H_0\} = \{q(k') : k \xrightarrow{a} k', k' \in H_0\}. \quad (15)$$

Equality in Equation 15 includes the possibility that both successor sets are empty. Consequently, a shared observed present is insufficient to license a common action unless both histories have corresponding observable successors for that action. The criterion concerns the selected transition relation and observation map. It is stronger than matching current observations, and weaker than requiring the underlying histories or material configurations to be identical. It is an application of standard behavioral-equivalence reasoning [28], adapted here to make provenance a separate component of the comparative warrant.

Proposition 6.2 (Observable continuation under a stable history domain). *Suppose H_0 contains every admissible successor reached from its histories by actions in A_0 , and Equation 15 holds for every pair of histories in H_0 sharing the same q -image. Then the observable description classes support a well-defined action-labelled transition relation. For every finite action sequence in A_0 and every two initial histories in H_0 with the same q -image, the sets of observable terminal descriptions attainable through that sequence coincide. The conclusion concerns attainable observable terminal sets under the declared actions, without identifying transition probabilities, durations, or causal histories.*

Proof. For each description z represented in H_0 and selected action a , define its observable successor set by choosing any representative h with $q(h) = z$. Equation 15 makes the definition independent of the representative. For a sequence of length zero, the terminal-description set contains the shared initial description. Assume equality of terminal-description sets after a sequence of length n . The set after one further selected action is the union of successor sets of all descriptions attained at length n . Representative independence and the induction hypothesis give the same union for either initial history. Closure of H_0 under selected actions permits every step. Induction proves the assertion. $\square$

Proposition 6.2 is a familiar quotient-transition argument. It establishes a precise sufficient condition for the prediction task described; it is not a new general theory of historical determination. In stochastic settings, equality of successor sets remains too weak to preserve probabilities, and equality of associated transition kernels or an appropriate probabilistic behavioral criterion would be needed. For histories subject to resource consumption, the history states must include enough information to determine the admissible firings. The conditions in Proposition 6.2 should thus be checked at the level corresponding to the intended inference.

6.4 A Counterexample from Divergent Formation Paths

A finite counterexample shows how current-description equality can conceal a difference in admissible future action. Begin with a common initial configuration s_0 . A *protect* operation produces a hidden condition s_p that preserves one unit of a required precursor; a *spend* operation produces s_d in which that precursor has been exhausted. Both conditions are displayed as $z = \text{ready}$ by the chosen present-state observation q . An *execute* operation is admissible from s_p and produces an outcome s_f , while no execute successor is available from s_d . The different formation paths are both admissible, and their endpoints share an observation. Nevertheless, their continuation sets differ for the selected action.

The crucial details are the omitted resource condition and the action under comparison. The observation *ready* may be entirely accurate for a narrow user-interface convention and still be insufficient for predicting execution. Recording the consumption event, or recording the remaining precursor directly, resolves the ambiguity for this particular task. The richer record need not retain every detail of the formation history. Its adequacy depends on whether the remaining information suffices for the relevant action and outcome claims.

This example should be distinguished from the consumptive-resource counterexample of Section 5. Section 5 examined several proposed operations drawing on a common endowment. The present example examines *two alternative histories leading to observationally equivalent present configurations* whose later capabilities differ. The problem is therefore historical sufficiency of the comparison record, even when only one future operation is being considered.

A corresponding practical error can occur when researchers infer a participant’s capacity from an output label without examining the procedures that granted access to the enabling infrastructure. The point is a comparative caution: different histories can yield the same observable descriptor, and a claim about subsequent capacity needs evidence concerning whichever historical dependencies remain operative. The actual existence and magnitude of such differences must be established for each empirical application.

6.5 Predictive Equivalence and Economical Provenance Records

Complete histories are often unavailable, costly to collect, and unnecessary for a defined prediction. The relevant objective may be to construct a *sufficient historical summary*: a representation that retains precisely the distinctions needed for selected future inferences. For a stochastic process with recorded past h , one ideal criterion is equality of conditional distributions over the future observables relevant to the study. It is stated in Equation 16 for histories on which the conditional probabilities are defined.

$$h \sim_{\text{pred}} k \iff \mathbb{P}(Y_{\text{future}} \in B \mid h) = \mathbb{P}(Y_{\text{future}} \in B \mid k) \text{ for all selected measurable } B. \quad (16)$$

Equation 16 expresses an established predictive-equivalence idea in the setting of a declared observational task. Computational mechanics provides a substantial formal tradition in which histories are grouped according to their predictive implications and minimal representations are investigated [30]. The present paper adopts that tradition as an intellectual resource and does not claim priority over it. Furthermore, a predictive statistic can be adequate for forecasting selected observables while omitting information needed to assess responsibility, legal provenance, interpretive legitimacy, or the validity of a counterfactual intervention.

The distinction between *predictive sufficiency* and *explanatory sufficiency* is consequently important. A summary can successfully forecast observable outcomes while collapsing histories with different causal structures. If a researcher intends to assess how a capacity was produced, which participant made a constitutive contribution, or how the system would respond to an intervention, the summary must preserve evidence pertinent to those questions. The relevant set of futures and interventions must be declared, and alternative hidden-state models should be considered. Forecasting accuracy alone licenses only the predictive claim evaluated.

History records can also be compressed for legitimate reasons unrelated to forecasting. A legal chain of custody may retain acts of transfer and authorization; a laboratory provenance record may retain calibration versions, instruments, and transformations; a linguistic inquiry may retain changes in usage communities and interpretive conventions. In each case, the admissible compression is relative to a purpose. More archival detail may improve reconstruction, but increasing detail without a clearly specified question may also increase noise, privacy exposure, and false confidence in the completeness of the record.

6.6 Comparison across Divergent Historical Regimes

Cross-system historical comparison introduces a difficulty beyond reconstructing one system's trajectory. The relevant histories may be described with different categories, recorded at different resolutions, or governed by different admissibility regimes. A mapping between current configurations may therefore require a second mapping between selected formation events and an explicit account of what each historical description omits. Such a mapping must respect the units and interpretive conventions of the particular inquiry. A manufacturing event, an ecological succession, and an institutional delegation can each be represented as a transformation while carrying different causal and normative implications.

A common comparative procedure can begin by identifying the target outcome or capacity whose origin is in question, selecting the relevant precursors and transformations in each system, and specifying which correspondence between these transformations is justified. For some questions a partial order over events is sufficient; for others durations, rates, or simultaneous interactions are essential. A formal relation between event records should preserve the dependencies on which the intended inference relies. Equivalent labels for events provide only a preliminary aid to interpretation. Evidence about enactment, execution, and material consequences is needed to establish substantive correspondence.

Historical comparison can also reveal distinct *counterfactual opportunities*. Two organizations may exhibit similar current practices while one has independent capacity to revise its protocol and the other remains dependent on external authorization. Two technological assemblages may deliver an identical output yet differ in their opportunities for repair because their components, interfaces, and supply conditions were generated through different arrangements. In such comparisons, the researcher should identify the precise mechanism connecting formation history to current feasible practice. The conclusions should remain confined to demonstrated dependencies rather than to any asserted universal ranking of autonomy or historical development.

Increasing-returns and lock-in accounts provide useful hypotheses where cumulative adoption or institutional reinforcement can be evidenced [3, 25]. Alternative mechanisms may include fixed material constraints, reversible optimization, common external shocks, measurement artifacts, or selective retention of records. An observed association between earlier adoption and later dominance can be compatible with several such processes. Historical sequencing and causal dependence consequently require different kinds of evidence. The comparative objective is to discriminate among these mechanisms, where feasible, while leaving unresolved cases explicitly open.

6.7 Provenance Documentation and Evidential Boundaries

A defensible historical comparison should make its provenance selection and its limits inspectable. Evidence about a past transformation may be missing, contested, or selectively preserved. Archives can reflect asymmetric recording power, measurement practices can change, and reconstruction algorithms can import assumptions about admissible sequences. A formal model cannot remove these difficulties by assigning a single path to an uncertain record. Where multiple histories remain compatible with the observations, the analysis may instead compare the conclusions supported across the admissible reconstructions.

Table 5 presents an operational checklist. It keeps distinct the existence of a record, the adequacy of a reconstructed history, and the validity of a causal or predictive conclusion.

Table 5: Evidence required for historically informed comparison of heterogeneous systems.

Comparative element	Evidence, test, or qualification to record
Current equivalence	Observation variables, resolution, matching rule, and conditions in which the present descriptions agree
Formation record	Event types, provenance sources, ordering assumptions, missing periods, and disputed accounts
Persistent dependency	Material, relational, institutional, or interpretive mechanism connecting prior formation to current capability
Future comparison	Selected actions, horizons, observable outcomes, and whether continued feasibility is modeled
Predictive sufficiency	Alternative histories sharing the present description, and tests for equal continuations or conditional distributions
Causal interpretation	Intervention assumptions, rival mechanisms, and evidence capable of discriminating historical dependence from confounding
Cross-system translation	Correspondence between event categories, regimes, and the historical distinctions lost in the mapping
Review and correction	Procedures for revising the history set, challenging provenance claims, and updating conclusions

The criteria in Table 5 are proposals for disciplined inquiry, rather than a validated generic scoring scale. Their application will vary with the kind of evidence available and with the ethical implications of collecting it. In human or institutional research, reconstructing histories may expose confidential participation, security-sensitive dependencies, or unequal opportunities to challenge records. Legitimate access, proportionality, and the ability of affected parties to contest an inference may therefore be conditions of responsible comparative practice.

Historical provenance tests whether present-state comparisons remain adequate in light of formation, retained constraints, and future possibilities. The analysis distinguishes descriptions that preserve selected continuations from those that conceal operative dependencies. Alongside the resource-sensitive criteria of Section 5, the present section provides a test of historical sufficiency. The following section examines reconstruction limits when observations support incompatible models.

7 Formal Limits of Cross-System Reconstruction

The reconstruction of heterogeneous generative systems presents an inverse problem: observations of configurations, outputs, and selected transformations must be related to underlying generative conditions that may remain partly inaccessible. The objective of this section is to determine which comparative assertions follow from such evidence when several explanatory models remain admissible. The argument proceeds from the definition of evidence-compatible reconstruction families to a criterion for identifying task-specific comparisons, finite countermodels, partial identification, and research designs capable of discriminating rival accounts. Its formal constructions are elementary set-theoretic and finite-model results. Their scholarly role lies in making explicit the conditions of inference across different systems; they introduce neither a universal inverse method nor a general solution to causal identification.

A comparison can be well defined while its evidential justification remains unsettled. For example, two systems may expose the same interface and yield the same recorded outputs under routine operating conditions, yet differ under an intervention that draws on unobserved resources or historical permissions. Structural correspondence, as formulated in Section 3, determines what a proposed map claims to preserve. Historical adequacy, as examined in Section 6, tests whether present descriptions retain the continuations relevant to a question. Reconstruction requires a further step: the investigator must establish how far the available observations constrain the models and correspondences from which these claims are inferred. The distinction is especially important where a description can be made precise long before the description has been empirically warranted.

7.1 Reconstruction as a Task-Relative Inverse Problem

Let A and B denote two candidate systems. A comparative reconstruction may concern their configuration spaces, operational rules, historical dependences, observational interfaces, and proposed maps between them. An observed record may contain selected input–output pairs, time-ordered events, measurements at a finite resolution, statements by participants, or institutional documents. These are different kinds of evidence. Their combination calls for declared correspondence assumptions, including the conditions under which measurements or terms from the two systems can be regarded as comparable.

The mathematical language of inverse problems supplies a limited but useful analogy. A forward model generates possible observations from stipulated dynamics or parameters. In the inverse direction, observations constrain the candidate models compatible with them. The methodological literature distinguishes forward specification, uncertainty, and the use of regularization or prior assumptions in reconstruction [18]. In G4, a forward model may include both a pair of systems and a proposed comparison map. Its output is a selected evidential record rather than a complete description of the relevant realities.

Represent a candidate reconstruction by $\theta = (M_A, M_B, f, \kappa)$, where M_A and M_B are system models, f denotes an appropriately typed partial correspondence, and κ records the regimes within which predictions and translations are admissible. The symbol θ names a candidate analytical construction; it does not presume that each component is independently identifiable. Let Θ be a declared family of candidates and $D : \Theta \rightarrow \mathcal{E}$ a forward evidence map into a set of possible observational records. A target comparison is represented by $\varphi : \Theta \rightarrow V$, where V may consist of truth values, numerical quantities, or structured claims. For a realized record e , the compatible candidates form the identification set in Equation 17.

$$\Theta(e) = \{\theta \in \Theta : D(\theta) = e\}. \quad (17)$$

Equation 17 assumes a fixed forward map and exact observational agreement. Both assumptions will be relaxed below. It immediately makes model selection assumptions visible: admissible configurations, correspondence functions, action repertoires, sensor resolutions, and historical alternatives are selected before the identification set can be calculated. A narrow Θ can create apparent certainty by excluding mechanisms whose relevance has not been independently examined. An unrestricted Θ may prevent informative conclusions because almost any finite record is compatible with many extensions. Responsible reconstruction therefore reports both the data and the principled limits of the candidate family.

7.2 Identification of a Comparative Claim

The ability to identify a claim about a pair of systems is weaker than the ability to reconstruct either system uniquely. Several models may disagree about hidden state descriptions while agreeing on the question for which the comparison was designed. Conversely, the identification of an observable output under one protocol can coexist with considerable uncertainty about behavior under another. The relevant object of identification is accordingly φ , rather than θ in its entirety.

Definition 7.1 (Task-relative comparative identification). For a fixed evidence map D , a target $\varphi : \Theta \rightarrow V$ is *identified at an evidence record e* when $\Theta(e)$ is nonempty and φ has a single value on $\Theta(e)$. The corresponding *identified-value set* is the image $\varphi[\Theta(e)]$. The target is *globally identified relative to D and Θ* when it is constant on each nonempty fibre of D . Identification here is a logical property of the stipulated family and evidence map; it is distinct from statistical estimation precision and from physical validation of the model family.

A useful elementary characterization records the dependence of identification on observational distinctions.

Proposition 7.2 (Factorization criterion for identified comparisons). *Let $D : \Theta \rightarrow \mathcal{E}$ and $\varphi : \Theta \rightarrow V$ be functions. The target φ is constant on every fibre of D if and only if there exists a function $\psi : D[\Theta] \rightarrow V$ satisfying Equation 18.*

$$\varphi = \psi \circ D. \quad (18)$$

Proof. If φ is constant on each fibre, define $\psi(e)$ as $\varphi(\theta)$ for any θ satisfying $D(\theta) = e$. Constancy makes this definition independent of the chosen representative, and Equation 18 follows. Conversely, the factorization assigns the same φ -value to all candidates sharing a D -value. $\square$

Proposition 7.2 is a standard factorization observation. It should be read as a conceptual test rather than as a new identification theorem. For a task-specific comparison to be recoverable from an observation record, the target may distinguish models only where the record itself distinguishes them. The proposition also clarifies an important limit of structural mapping: even an exact formal correspondence between stipulated models supplies an identified empirical comparison only when competing models compatible with the evidence agree on the relevant mapped property. Reconstruction is thus constrained by the observational partition of the candidate family.

In practice the target φ may be the truth of a selected generative capability, a relation of joint feasibility, a numerical conservation quantity, or the availability of a history-sensitive continuation. These claims can have different identification statuses under the same evidence. A recorded equality in output rates can identify the comparison of those recorded rates while leaving the correspondence of underlying mechanisms unresolved. This is one reason to document comparison tasks separately from the stronger explanatory interpretations sometimes attached to their results.

7.3 Countermodels with Indistinguishable Observations

A finite example exposes the limits of passive observations. Suppose two systems A and B each have an observable resting state 0 and an observable marker state 1. Under a routine action r , both remain at 0. Under a separately specified probe action u , a hidden generative rule may make the marker attainable. The available data consist of repeated observations of 0 during executions of r , together with the fact that the selected interface reports only 0 or 1.

Consider two deterministic candidate reconstructions. In model θ_+ , system A reaches 1 after u and system B remains at 0; in model θ_- , the assignments are reversed. Both reproduce the entire selected routine record. Yet the comparative claim φ that “ A admits the marker-generating probe and B does not” is true in θ_+ and false in θ_- . Consequently, the observational fibre contains both truth values. The current observations identify the shared routine behavior and leave the comparative probe claim unresolved.

This counterexample establishes a logical insufficiency: a common observational trace under a restricted protocol gives no general entitlement to infer comparative capabilities under unobserved actions. The example involves two fully specified deterministic models; introducing measurement noise or hidden histories would provide further sources of ambiguity. Conversely, an appropriately executed and adequately measured probe u could distinguish the stated candidates, subject to the legitimacy and feasibility of performing it. Whether a proposed probe is a permissible operation in a biological, institutional, or social system is an independent matter of ethics and governance. The logical availability of an experiment supplies neither an obligation nor an unrestricted authorization to conduct it.

The example also highlights the importance of cross-system interpretation. One might observe symbols designated “0” and “1” on both interfaces while the markers refer to different kinds of event. The comparison requires independent justification that the observable categories express the same target property or a sufficiently specified translation of it. If the mapping itself is uncertain, several correspondence assignments may remain compatible with the same token-level record. Section 3 gives the local mapping criteria; reconstruction analysis must now include uncertainty in the candidate maps as well as in the systems they relate.

7.4 Alternative Mappings and Structural Underdetermination

A comparative model can fail to be identified even when the separate systems have been measured with considerable precision. Suppose the first system distinguishes two generative events a_1 and a_2 , while

the second reports only an aggregated event category b . Both correspondences $f(a_1) = b$ and $g(a_2) = b$ may preserve all currently recorded single-event outputs. Their implications can diverge for a joint prerequisite involving a_1 but excluding a_2 . The available evidence can then identify a broad output analogy while leaving the event-level translation underdetermined.

The mathematical source of this limitation is familiar: a many-to-one observation may retain selected invariants while losing information about individual representatives. Section 3 already states an exact relation-quotient criterion. The present point concerns inference about whether a proposed quotient criterion actually holds for the world under investigation. Observed agreement on a subset of relation instances leaves open the behavior of unseen instances. Testing the quotient requirement may therefore demand additional joint-input experiments, independent measurement of hidden roles, or evidence that the selected domain exhausts the relevant possibilities.

For some comparisons, a family of admissible partial correspondences is the appropriate result. This family may contain mappings that preserve a numeric invariant while disagreeing about causal interpretation, or mappings that preserve local event sequences while failing to preserve a conservation constraint. Where a translation requires selecting an interpretation of a human concept, the candidate family may include contested semantic and institutional assumptions. A mathematically precise translation table can be useful without making those assumptions disappear. Its validity needs to be evaluated within the linguistic, practical, and evidential circumstances of the particular comparison.

The formal literature on latent-structure models provides an illuminating, bounded analogy. Identifiability of model parameters can depend on the number and arrangement of observed variables and on specific independence assumptions [2]. That work establishes results for particular statistical model families. It supports the general methodological lesson that additional observation channels and structural assumptions can change identifiability; it supplies no automatic identifiability theorem for G4's heterogeneous systems. For cross-system inquiry, the corresponding research obligation is to declare the observation channels, their transformations, and the assumptions that connect their outputs.

7.5 Partial Identification under Incomplete Evidence

Exact identification is sometimes an unnecessarily strong objective. A comparison can be informative when the evidence establishes a set of possible values, a one-sided bound, or a collection of claims shared across competing models. Set-valued inference has an established methodological literature, particularly in work on partial identification [21]. G4 uses this distinction in a task-relative manner: the investigator can state what remains true under every explicitly admissible reconstruction while recording which alternatives remain open.

For a nonempty family $\Theta(e)$ and a numeric target $\varphi : \Theta \rightarrow \mathbb{R}$, define the interval enclosure in Equation 19.

$$I_\varphi(e) = \left[\inf_{\theta \in \Theta(e)} \varphi(\theta), \sup_{\theta \in \Theta(e)} \varphi(\theta) \right]. \quad (19)$$

Equation 19 allows extended endpoints when the admissible values are unbounded. The interval may contain values unattained by any compatible model; the identified-value set $\varphi[\Theta(e)]$ remains the sharper object. An interval is appropriate only when an ordering and scale for the target have been justified. For categorical or normative comparisons, the explicit set of admissible conclusions is more faithful than a numerical surrogate.

A boolean comparative claim P admits two complementary assessments. It is *warranted relative to the candidate family* when it holds under all $\theta \in \Theta(e)$; it is *possible relative to the family* when it holds under at least one such θ . Under these definitions a warranted claim is possible whenever the candidate family is nonempty. A failed universal test leaves a claim unsettled until a countermodel is examined; it does not establish a physically realized counterexample. Likewise, a model that permits an outcome does not show that the outcome will occur in the investigated system.

Proposition 7.3 (Monotonic refinement of identified claims). *Suppose evidence e' refines evidence e within the same target and candidate family, so that the nonempty compatibility sets satisfy $\Theta(e') \subseteq \Theta(e)$.*

Then the identified-value set for any target φ cannot enlarge. Every boolean claim warranted under e remains warranted under e' , while every claim possible under e' was already possible under e . For a real-valued target, its lower bound can only increase and its upper bound can only decrease.

Proof. Inclusion of compatibility sets implies inclusion of their images under φ . Universal claims continue to hold on subsets, existentially satisfied claims on subsets were satisfied on the original set, and the infimum and supremum inequalities follow from set inclusion. $\square$

The conclusion of Proposition 7.3 is conditional on a fixed, correctly specified family. Revising the candidate assumptions, discovering an omitted mechanism, or correcting a mistranslation can enlarge or replace the admissible set. Scientific learning can therefore invalidate an earlier warrant without contradicting the proposition: the earlier and revised analyses differ in their model assumptions. This is a reason to preserve a revision history for comparative claims rather than treating identified bounds as permanent properties of an observed system.

7.6 Model-Family Sensitivity and Conditional Identification

Identification within a stipulated model family also depends on the adequacy of that family's boundaries. A diagnostic distinction is required between the arrival of new observations under a fixed family and the admission of additional models under an unchanged record. The first operation can eliminate compatible candidates as described by Proposition 7.3; the second can restore possibilities that the original analysis had excluded by assumption. The two operations have different inferential directions and require different kinds of justification. Manski's treatment of partial identification provides an established methodological point of reference for the dependence of identified conclusions on assumptions [21].

Consider nested families $\Theta_0 \subseteq \Theta_1$, with the same evidence map D and target φ defined on the larger family. For an observed record e , compatibility under the smaller family is the intersection $\Theta_0(e) = \Theta_0 \cap \Theta_1(e)$. Consequently the identified-value sets satisfy the inclusion in Equation 20.

$$\varphi[\Theta_0(e)] \subseteq \varphi[\Theta_1(e)]. \quad (20)$$

The inclusion is a direct set-theoretic consequence, conditional on a common target, common observation rule, and nonempty compatible sets. It does not imply that the larger family is physically more accurate. Its analytical role is to expose how an apparently exact conclusion may rest on excluding mechanisms that remain observationally admissible.

For a finite example, take two candidate systems θ_0 and θ_1 with the same routine observation $D(\theta_0) = D(\theta_1) = e$. Suppose a specified intervention produces the target outcomes $\varphi(\theta_0) = 1$ and $\varphi(\theta_1) = 0$. Under the declared family $\Theta_0 = \{\theta_0\}$, the target is point-identified at the value 1. Under $\Theta_1 = \{\theta_0, \theta_1\}$, the identified-value set is $\{0, 1\}$. The observational record has remained unchanged; the difference arises from an expanded space of admissible explanations. This example is constructed to demonstrate logical sensitivity, rather than to estimate the frequency of omitted mechanisms in actual research.

A practical sensitivity audit should therefore record at least one justified enlargement or alteration of the candidate family, examine whether the comparison survives the added countermodels, and specify the substantive reasons for including or excluding each alternative. The family should be altered through plausible mechanistic, historical, or interpretive alternatives, rather than arbitrary constructions unrelated to the investigation. Where a purported invariant survives a sufficiently varied set of independently justified candidates, the researcher can report that conditional robustness along with its remaining limits. Where its value changes, the comparison remains appropriately bounded by the narrower set of assumptions.

The distinction is especially important when researchers compare systems that disclose different kinds of evidence. Richer documentation in one system may reveal alternatives that remain hidden in another. Such asymmetry should enter the record as a difference in evidential coverage, while the intended comparative property remains specified separately. Section 11 incorporates model-family sensitivity into

the proposed evaluation procedure; a claim identified in a convenient reconstruction family should not be mistaken for a conclusion guaranteed by the observed systems themselves.

7.7 Uncertainty in Observation and Model Adequacy

Finite resolution, missing observations, and noise complicate the evidence map D . A deterministic equality criterion may falsely exclude models because a measured value differs slightly from its idealized prediction. One response is to specify a set of admissible observed records for each model. Let $\mathcal{D}(\theta) \subseteq \mathcal{E}$ contain the records permitted by the model and its declared error conventions. Compatibility then means $e \in \mathcal{D}(\theta)$. Equivalently, a numeric observation y and model prediction $g(\theta)$ can be related by a declared tolerance, a stochastic measurement model, or an explicit uncertainty set. These alternatives serve different scientific purposes and should be chosen on substantive grounds.

An error tolerance alone does not identify the correct mechanism. If two models yield predictions within the tolerance for the available observations, the comparison can remain underidentified. Bayesian weighting can assign relative plausibilities to models, but posterior conclusions depend on likelihoods and prior choices. Frequentist uncertainty statements depend on sampling and coverage assumptions. Set identification deliberately separates these modeling commitments from the logical range of conclusions compatible with stated evidence. Each approach can be appropriate for a different question, and their inferential outputs should be reported under their own assumptions rather than translated into a single undifferentiated label of confidence.

The distinction between structural and practical identifiability is also relevant. A parameter may be identifiable from idealized, unlimited observations of a correctly specified model yet remain poorly constrained by the available finite, noisy sample. Conversely, reliable prediction of a selected observable can be possible despite nonidentification of latent mechanisms. Statistical inverse-problem methodology provides established techniques for examining these difficulties [18]. G4 requires the investigator to state which kind of inferential object is being assessed: an observable comparison, an effective predictive model, or an explanatory claim about hidden generative conditions.

These distinctions prevent a further category error involving semantic comparison. Some uncertainties concern missing facts about a system; others concern whether different communities or theories are using the same categories at all. Collecting more data under an unchanged instrument can reduce the former while leaving the latter intact. A redesigned comparative protocol may require negotiation of concepts, independent translation checks, or separate reporting of incompatible but internally coherent interpretations. Those practices are developed philosophically in the next section and require their own forms of evidential accountability.

7.8 Discriminating Designs and Comparative Research Practice

A useful reconstruction programme should connect its admitted uncertainty to possible discriminating evidence. For any target comparison φ , identify two compatible models that yield different values. Then ask which permissible observation or intervention, if any, would yield different predicted records under those models. A fully discriminating test for a finite model family would separate every pair of candidates that disagree on φ . In actual research, observation distributions can overlap, actions can alter the systems, access may be restricted, and the candidate family can be incomplete. The design target should therefore be a stated reduction in comparative ambiguity under a transparent protocol, with failure to distinguish models retained as an informative outcome.

The task can be organized into a sequence of explicit decisions. The researcher identifies the comparative claim, the supported domains, the observation or intervention repertoire, and the allowable model families. Proposed translations between descriptions are recorded as candidate maps. Each model is checked for compatibility with existing evidence. The analyst then reports identified claims, possible alternatives, and countermodels. Further data collection is prioritized according to its expected capacity to discriminate materially different conclusions, subject to consent, safety, resource feasibility, and

Table 6: Diagnostic distinctions for assessing reconstruction and identification across heterogeneous systems.

Source of ambiguity	Evidence requirement and warranted reporting practice
Restricted action set	State the observed interventions; retain countermodels differing under unobserved admissible actions
Measurement coarsening	Record resolution and aggregation; assess whether relevant transformations or prerequisites survive observation
Alternative correspondence	Declare candidate mappings and tests of preserved or reflected relations
Historical omission	Compare alternative formation histories and the selected continuations they permit
Latent dynamics	Distinguish identifiable outputs from unidentifiable parameters or mechanisms
Noise and finite samples	Specify error models, tolerances, sample support, and sensitivity to modeling assumptions
Model-family choice	Record excluded mechanisms, assumptions, and the effects of expanding the family
Discriminating observations	Identify ethically permissible tests that predict different records under competing conclusions

institutional authorization. These steps constitute a provisional research workflow, not an empirically validated universal comparison procedure.

Table 6 separates the principal sources of underidentification and their appropriate evidential responses. The table is deliberately diagnostic: it cannot substitute for domain-specific measurement theory, historical scholarship, or causal research design.

The diagnostics in Table 6 also constrain claims about invariance. A property common to all models supported by a bounded observational record is invariant within that candidate family; whether it is invariant in the investigated systems requires further justification of the family, measurement procedures, and cross-system mappings. Formal identification, evidential confidence, causal explanation, and ontological interpretation can therefore be evaluated separately, even when they concern the same comparison.

A heterogeneous comparison can remain scientifically productive while its central claims are only partially identified. Reporting compatible alternatives makes visible what existing evidence can sustain and where additional measurement, intervention, interpretive work, or conceptual revision may matter. The next section turns to reciprocal interpretation, where comparative categories themselves can be transformed through interaction among differently situated participants.

8 Reciprocal Interpretation and Epistemic Translation

The comparison of heterogeneous generative systems may involve participants who classify events, articulate evidence, and revise categories through interaction. Such encounters introduce an interpretive dimension beyond the correspondence of preassigned structures examined in Sections 3–7. The objective of this section is to explain how a comparative relation can be jointly articulated when the categories used to identify its terms are themselves open to negotiation. The discussion distinguishes linguistic interpretation, operational translation, reciprocal testing, and institutional authority; it then gives two deliberately limited formal criteria for evaluating selected translations and develops an evidence protocol for maintaining productive disagreement. The method is philosophical and comparative, with elementary mathematics used to delimit particular claims of preservation. Neither convergence of vocabularies nor agreement between interlocutors is assumed in advance.

Interpretation is a relation among situated inquirers and their environments, mediated by practices of reference, inference, action, and evaluation. A scientific instrument, a legal institution, a software agent, a biological community, and a human interlocutor can each participate in an inquiry in different

capacities. Some provide records or undergo interventions; others can articulate, criticize, or authorize interpretations. The word *reciprocal* applies most directly where both sides can answer, correct, or revise a proposal through some accessible channel. For other cases, the investigator may pursue a carefully qualified analogue through repeated observations, alternative models, and independent criticism. A functioning interface alone supplies insufficient grounds for attributing reciprocal understanding, consciousness, consent, or moral agency.

8.1 Interpretive Coordinates and Situated Meaning

A comparison requires decisions about the units that are being compared. These may be words, gestures, measurement categories, state descriptions, action types, or relations between events. For an interlocutor embedded in a particular practice, the significance of a term may depend on characteristic uses and on the consequences of accepting or rejecting its application. A second interlocutor may distinguish a different set of cases, even where familiar words or the same symbols are used. For example, a laboratory's category of successful output may be defined by an assay threshold; a clinical service may also require reproducibility, patient benefit, and acceptable risk. The resulting classifications can overlap extensively while answering different questions.

Davidson's account of radical interpretation locates interpretation within the joint problem of attributing meaning and belief, while criticizing the idea of wholly insulated conceptual schemes [7]. Discussion of scientific incommensurability, by contrast, highlights cases in which historical or disciplinary taxonomies resist complete lexical substitution; a qualified comparison can remain possible through learning and situated reconstruction [24]. The present inquiry treats these as distinct philosophical contributions. It does not infer from their disagreement either a universal translation mechanism or the impossibility of communication across heterogeneous systems.

A useful comparative record therefore specifies at least four coordinates: the source practice in which an expression operates; the target practice in which its proposed counterpart operates; a declared question motivating the comparison; and the occasions on which participants or observers can challenge its application. Two labels may remain aligned across a small repertoire of actions and diverge under unfamiliar circumstances. The scope of the repertoire is part of the warrant for comparison, rather than an incidental limitation of notation. Section 7 addressed the identification of a target from a fixed observational record. Here the observational categories themselves may be revised during inquiry, changing the target and the admissible reconstruction family together.

8.2 Negotiation of Comparative Categories

Category negotiation occurs when participants examine the extension, permissible inferences, and consequences of using an expression as a counterpart of another. Consider a coastal monitoring collaboration in which an engineering team treats an area as *accessible* whenever an instrument can physically reach it, while local stewards distinguish physical reachability from an authorized opportunity to enter. Both descriptions may be accurate within their institutional settings. For the cooperative question of whether a sampling operation should proceed, the term requires an agreed operational interpretation that specifies the physical and authorization conditions separately. The example illustrates a possible analytical situation; it is neither a report of an observed community nor a claim that its participants hold any particular beliefs.

Negotiation may proceed through examples, counterexamples, demonstrations, revised measurement procedures, and records of disagreement. Parties can identify a common operational domain without replacing their complete conceptual schemes. A stable result might be a carefully delimited translation table, an annotated decision rule, a set of matched observations, or the decision to retain two terms and describe their relation explicitly. The success criterion depends on the inquiry: faithful description, prediction, coordination, or an ethically accountable decision can impose different demands on the translation.

A temporary comparative vocabulary can be useful when its revision conditions are recorded. Its categories should identify the episodes used in construction, the kinds of new cases likely to challenge them, and the parties who can request correction. Such a vocabulary is best understood as an accountable working interface. It is neither a final common language nor evidence that all significant disagreements have been resolved. The provisional character of categories is compatible with precise local specifications and reproducible tests.

8.3 Reciprocal Partial Maps and Round-Trip Adequacy

The partial mappings of Section 3 can be enriched by allowing a proposed translation to be checked in both directions. Let S_A and S_B be sets of selected expressions, classifications, or operational descriptions. Let $D_A \subseteq S_A$ and $D_B \subseteq S_B$ be comparison domains. A partial map $f : D_A \rightarrow D_B$ carries selected terms of practice A into practice B , and g is defined on a domain including $f[D_A]$ with values in S_A . Equivalence relations $\sim_A$ and $\sim_B$ express task-relevant, independently specified agreements of use; they need not encode identity of intrinsic meaning. Equation 21 records an optional round-trip condition.

$$g(f(a)) \sim_A a \quad (a \in D_A). \quad (21)$$

Equation 21 requires more than a shared label, since the returned interpretation must satisfy the declared comparison relation in the source practice. Yet the test is local: the map may be undefined on expressions outside D_A , and the relation $\sim_A$ may disregard distinctions that become decisive for another task. For example, two assay outcomes may be equivalent for initial screening and distinct for treatment selection. Reciprocal adequacy must therefore record both its domain and the particular equivalence criteria used to judge a successful return.

Proposition 8.1 (Restricted discrimination from reciprocal translation). *Suppose $f : D_A \rightarrow D_B$ respects $\sim_A$ and $\sim_B$ on its declared domain, and $g : f[D_A] \rightarrow S_A$ respects $\sim_B$ and $\sim_A$ on the image of f . If Equation 21 holds for every $a \in D_A$, then f reflects equivalence on D_A : whenever $f(a) \sim_B f(a')$, one has $a \sim_A a'$. Consequently, f induces an injective map between the corresponding classes in $D_A/\sim_A$ and $D_B/\sim_B$. It need not be surjective onto all target equivalence classes and therefore establishes neither a global inverse nor equivalence of complete semantic systems.*

Proof. If $f(a) \sim_B f(a')$, preservation by g yields $g(f(a)) \sim_A g(f(a'))$. The two round-trip conditions yield $a \sim_A g(f(a))$ and $g(f(a')) \sim_A a'$. Transitivity proves $a \sim_A a'$. Preservation by f makes the map on classes well-defined; reflection makes it injective. $\square$

Proposition 8.1 is an elementary fact about equivalence relations and partial functions. Its use in the present inquiry is diagnostic: a reciprocal test can detect the collapse of distinctions that the source practice regards as relevant. The result supplies neither surjectivity onto all target categories nor a guarantee that source and target equivalences capture the same historical experience. Successful round-trip performance can also depend on negotiated labels that already discard a disputed distinction. If the domain is repeatedly reduced to cases on which agreement is easiest, apparent translation success may become increasingly uninformative for the original question.

8.4 Shared Probes and Task-Relative Preservation

Some comparisons can be assessed through a finite or otherwise specified family of joint probes. A probe may be a controlled physical intervention, a classificatory vignette, a request for an explanation, or an agreed calculation. Its interpretation includes an input protocol, a record of the response, and a rule for evaluating that response. Agreement on a numerical display is insufficient if two instruments use incompatible units, calibration procedures, or operating conditions. In human communication, the interpretive task may also require recording why different respondents regard a probe as relevant.

Let T be a declared collection of probes and let W_t be a calibrated outcome space for each $t \in T$. For $i \in \{A, B\}$, a selected signature map σ_i associates each admissible description with its probe results

in $W = \prod_{t \in T} W_t$. Whenever an outcome is inaccessible or uncertain, this status should be represented explicitly; the following elementary proposition assumes well-defined exact signatures. A target result h_i may be factored through these signatures when the chosen task depends only on the declared probes.

Proposition 8.2 (Preservation of probe-factorable comparisons). *Let $f : D_A \rightarrow D_B$ be a partial comparison, and let $\sigma_A : D_A \rightarrow W$ and $\sigma_B : D_B \rightarrow W$ be signatures for a shared calibrated probe family. Suppose $\sigma_B \circ f = \sigma_A$ on D_A . If a selected comparison result is represented by $h_A = H \circ \sigma_A$ and $h_B = H \circ \sigma_B$ for the same $H : W \rightarrow V$, then $h_B(f(a)) = h_A(a)$ for every $a \in D_A$.*

Proof. For $a \in D_A$, substitution gives $h_B(f(a)) = H(\sigma_B(f(a))) = H(\sigma_A(a)) = h_A(a)$. $\square$

Proposition 8.2 is a direct factorization identity. It guarantees preservation only for results that genuinely factor through the shared signature, subject to the accuracy of the observational and calibration assumptions. It gives a precise account of why local success can be useful: investigators may agree on the response to a clinically or technically relevant set of tests without having equivalent theories of the system. Conversely, a previously untested outcome may distinguish systems with identical established signatures.

A finite illustration makes the boundary visible. Suppose two control systems yield the same response under a readiness check and an initial action probe. One system can repeat the action after consuming a resource, while the other requires replenishment. A proposed equivalence based only on the first two probes is warranted for the specified short operation. It provides inadequate evidence for repeated operation. This is a constructed counterexample, rather than a finding about any deployed system. Its significance is methodological: a record of agreement should travel with the class of questions for which agreement has been established. Extending that record to a new class requires additional probes, assumptions, or both.

8.5 Divergent Epistemic Standards and Disagreement

Reciprocal interpretation can be sustained without an immediate agreement about how a disputed claim should be evaluated. Different practices may attach different weights to replication, testimony, historical experience, predictive accuracy, institutional authorization, or direct intervention. Their differences sometimes concern the quality of evidence available, sometimes the relevance of that evidence to a chosen conclusion, and sometimes the stakes attached to possible error. These distinctions are important for deciding what a comparative inquiry can reasonably settle.

Longino's account of scientific knowledge identifies social interaction and the role of values in the relevance assigned to evidence as central problems of scientific objectivity [19]. Within G4, this supports attention to how evidential standards are made explicit and rendered open to criticism. It does not entail that any negotiated convention thereby counts as scientifically adequate. Agreements that suppress counterevidence, disregard measurement uncertainty, or make criticism practically impossible would still require independent assessment. Equally, a well-designed technical probe may answer a narrow predictive question while leaving an ethical or historical dispute unresolved.

An interpretive encounter can therefore produce several legitimate outcomes. Participants may adopt one jointly calibrated category, maintain parallel classifications with a stated translation relation, agree on which new evidence would settle a question, or record a principled disagreement. The status of an unresolved category matters. Treating silence as assent or converting an unanswered question into an affirmative classification can change the evidential meaning of a record. The comparative archive should distinguish a proposed translation, an endorsed translation, a contested translation, and a translation awaiting decisive evidence.

The possibility of disagreement also motivates a distinction between *understanding a claim* and *accepting it*. An investigator may understand another participant's criteria sufficiently to predict their classifications while rejecting the relevance of those criteria for the shared task. Conversely, an institution may formally accept a translated classification while lacking the evidence needed to understand the practice that gave it meaning. Reciprocal interpretation benefits from recording these differences

rather than compressing them into a binary measure of agreement. This approach preserves space for alternative theories and respects Section 7's emphasis on rival admissible reconstructions.

8.6 Interpretive Authority and Unequal Participation

The distribution of interpretive authority affects which categories enter a comparative record and whose objections produce revisions. In a research collaboration, access to instruments, the ability to publish or archive records, control of translation interfaces, and authority to define standards of validation can be unevenly allocated. An apparently reciprocal exchange may take place inside an institution that recognizes the questions of one party and classifies the other's contributions solely as data. Formal opportunities to offer corrections may consequently differ from effective opportunities to influence the working vocabulary.

Fricker's analysis of epistemic injustice describes ways in which credibility deficits and shared interpretive resources can be shaped by social power [12]. G4 uses this literature as a reason to examine the conditions of interpretive participation, while retaining its separate analytical object: the warrant for a cross-system comparison. Concepts developed to describe injustice among human knowers should be transferred cautiously to artificial or other nonhuman systems. A system's production of legible outputs does not by itself establish the kind of social standing or experience presupposed by those normative theories. The possibility of unequal treatment of human participants, by contrast, can be assessed through concrete procedural and institutional evidence.

The normative issue has an epistemic counterpart. If one party controls the available probe family, the translation may preserve exactly the distinctions favored by that party. Repeated success under those probes may then be mistaken for an independently established common ontology. A more credible procedure specifies who selected the probes, the alternatives that were excluded, the costs of answering, and whether an affected participant can challenge a category or request additional evidence. These considerations anticipate Section 10's planned examination of authority over comparative categories. The present section focuses on their role in maintaining a usable and revisable translation.

Reciprocity should also be evaluated in relation to material conditions. A participant may be able to formulate a correction yet lack time, resources, access, or a safe opportunity to have it considered. An accessible review procedure and a recorded response to criticism are therefore relevant evidence for the claimed reciprocity of a comparative process. They are neither sufficient for broader justice nor substitutes for subject-matter expertise. Their value lies in making the production of categories inspectable and allowing disagreements to remain visible in the institutional record.

8.7 Comparative Episodes and an Evaluation Protocol

A reciprocal translation can be documented as a series of bounded comparative episodes. Each episode has an initiating question, a declared scope of expressions or actions, a proposal for correspondence, a selected probe family, and records of responses and corrections. Episodes need not share a universal clock or a single interpretation of temporal sequence. When events are reconstructed across different records, the relevant ordering assumptions and admissible synchronizations should be stated. This permits the inquiry to include historical documents, physical measurements, and contemporary dialogue without treating them as interchangeable kinds of evidence.

Table 7 presents a provisional evaluation protocol. It combines formal and interpretive criteria while specifying the consequence of a failure in a limited test. The table is intended for comparative case design, not as a validated scoring instrument. Each item may be operationalized differently across disciplines, and some kinds of cases do not permit direct reciprocity or deliberate probing.

Consider a hypothetical interorganizational monitoring project in which two institutions both report a parameter called *recovery*. One uses the expression for return of an instrument to a specified operating range; the other uses it for restoration of a living system's functions following disturbance. A translation that equates the labels can support a very limited comparison of operational records if both sides identify the associated measurements. It cannot, on that basis, establish equivalence of ecological recovery and

Table 7: Evidence and revision conditions for reciprocal epistemic translation.

Comparative element	Evidence to retain	Interpretive consequence
Declared translation domain	Source uses, target uses, excluded cases	Limits the reach of a correspondence claim
Calibration and probes	Operating conditions, units, failed and successful tests	Identifies the selected results that may transfer
Round-trip correction	Participants' proposed translations and returned interpretations	Reveals distinctions lost under a partial map
Alternative categories	Rejected proposals, unresolved terms, counterexamples	Preserves live explanatory and semantic rivals
Participation and authority	Access to challenges, decision procedures, recorded replies	Qualifies claims of reciprocal agreement
Revision triggers	New cases, altered regimes, changed institutional permissions	Specifies when a comparison requires re-examination

equipment readiness. A more careful episode would preserve both local definitions and introduce a jointly specified statement about which measurements can be compared, for which research purpose, and with what uncertainty.

A different hypothetical case involves dialogue between two professional groups who make incompatible judgments about an AI-assisted clinical recommendation. One evaluates model calibration against a retrospective dataset; the other evaluates the feasibility of acting on the recommendation in a particular service setting. Their disagreement may persist even when they agree on the numerical prediction. The relevant cross-system inquiry concerns whether the criterion of successful action can be related to predictive performance under explicitly specified institutional conditions. The answer may involve additional implementation evidence rather than a single translation of the word *success*. The example is illustrative and entails no recommendation for clinical deployment.

A useful evaluation record would expose both local standards and record how they influence the inferences drawn from shared observations. The procedure would also identify evidence that could invalidate the translation: a new case outside the original domain, an outcome dependent on a previously ignored resource, a disputed calibration, or a correction that changes the operative category. This makes interpretive success assessable as a bounded achievement, with pathways to further inquiry. It also avoids the inference that a negotiated category acquires general legitimacy solely because an exchange produced agreement.

8.8 Continuity of Comparison under Interpretive Revision

The formation of a comparative category changes the conditions for subsequent comparison. When a new distinction is introduced, the permissible correspondences between the two systems can be refined, but the relevant domains and target questions may change as well. As Section 7 showed for a fixed model family, adding evidence can shrink the set of candidate reconstructions. In reciprocal interpretation, category formation can also create new questions or new descriptions, so successive stages need not be related by simple set inclusion. An audit trail should record both the informational revision and the change in the object being compared.

This suggests a restrained account of *interpretive continuity*. An inquiry can remain continuous through disagreement when participants retain an accessible record of prior criteria, can explain revisions, and can identify what survives from earlier results. Continuity may attach to a practice of criticism and correction even where a particular translation is abandoned. The research programme can study such continuity without assuming that the parties converge toward a single fixed vocabulary, share identical values, or occupy equal institutional positions.

Section 8 has therefore supplied two modest formal criteria and a broader philosophical discipline for reciprocal interpretation. Round-trip conditions can protect selected distinctions against collapse,

while calibrated probes can preserve claims that genuinely factor through shared evidence. Both results are restricted by their domains and assumptions. The interpretation of what counts as a suitable domain, relevant equivalence, or admissible probe remains an object of situated inquiry and contestation. Section 9 will use this framework in comparative case designs involving natural and artificial systems; Section 10 will examine the institutional authority through which comparative categories are established and revised.

9 Comparative Cases across Natural and Artificial Systems

The formal distinctions developed in the preceding sections acquire methodological significance when an actual comparison requires choices about measurement, intervention, and inference. This section constructs four bounded comparative exercises involving synthetic biological regulation and computation, ecological food webs and material flow, multilingual grammatical annotation, and the reconstruction of histories and shared constraints. Its objective is to identify what selected correspondences can warrant in each setting, which competing explanations remain available, and what evidence would distinguish them. The cases are research designs anchored in published studies and publicly described infrastructures. They constitute neither a reanalysis of those sources' data nor empirical confirmation of a common generative ontology. The discussion proceeds by specifying the task, recording the admissible transformation and its potential remainder, and identifying a discriminating observation or intervention. A final protocol translates those choices into a reproducible comparative record.

9.1 Selection of Cases and Evidential Status

A cross-domain comparison is informative when it puts a proposed correspondence under pressure. Choosing examples solely because their diagrams appear similar would favor a result before the comparison begins. The cases considered here were selected to separate kinds of evidence that can coincide superficially: recurring output patterns, graph connectivity, annotation labels, and identical present observations. Each also offers a plausible established specialist baseline against which G4's proposed cross-relational account can be assessed. A successful specialist explanation is treated as a substantive result, even if the G4 vocabulary adds no predictive improvement.

Three evidential levels will be maintained. *Documented source systems* are the published experimental, ecological, or linguistic projects identified by citations. *Constructed comparison models* are simplified mathematical or interpretive objects introduced here to expose a distinction. *Proposed empirical tests* identify data and procedures that would be required to evaluate a claim of preservation. Source-level documentation may establish that a scientific object exists and clarify how it has been investigated; it does not establish the proposed cross-system comparison. Likewise, a correct result for a toy transition system supplies a counterexample or conditional illustration under its assumptions, with any physical interpretation requiring independent work.

A comparative case should specify at least an object or population of objects, a research question, a pair of admissible descriptions, the scope of a proposed mapping, the observation procedure, and the outcome that would count as a failed comparison. The proposal to use this record is organizational: it follows the task-relative comparative adequacy of Section 2, the partial relations of Section 3, and the identifiability restrictions of Section 7. It supplies no universal measure through which biological, ecological, linguistic, and computational phenomena can be ranked.

9.2 Synthetic Regulatory Oscillations and Computational Cycles

Elowitz and Leibler reported the construction of a synthetic genetic circuit of three transcriptional repressors in *Escherichia coli*, termed the repressilator, with a fluorescent reporter used to observe oscillatory behavior [9]. This experimental achievement furnishes a concrete biological setting in which a cyclic

organization of influences and a recurring observable output are both relevant. A corresponding computational object might be a cyclic finite-state controller whose internal states succeed one another in a fixed order. Each system admits a schematic directed cycle. Such a diagram is an explicitly selected abstraction of regulatory influence in the biological case and an operational transition representation in the computational case. The different meanings of an edge are part of the comparison itself.

A preliminary mapping might associate three coarse regulatory phases with three controller states. The task could ask whether the chosen diagram preserves the *existence of a directed cycle* under an agreed assignment of coarse phases. With a precisely specified finite-state abstraction, a graph isomorphism can answer that limited question. Stronger questions concern residence-time distributions, sensitivity to perturbations, stochastic variability, and the response of the system after an interruption. These require further information about kinetic laws, measurement thresholds, and the correspondence between biological and computational states. The biological observations summarized in the published report do not themselves establish such a mapping or numerical equivalence to a finite-state controller.

A particularly simple constructed countermodel clarifies the issue. Consider two continuous-time processes on an identical directed three-state cycle, where the only permitted transition from each state has rate $\lambda > 0$. The support graph is the same for all positive rates. If τ denotes the time to the next transition, the stipulated exponential waiting-time model gives Equation 22.

$$\Pr_{\lambda}(\tau > t) = \exp(-\lambda t), \quad t \geq 0. \quad (22)$$

The right-hand side of Equation 22 differs between rates even though reachability on the three states remains unchanged. For example, at $t = 1$ it is approximately 0.368 for $\lambda = 1$ and 0.050 for $\lambda = 3$. These are exact predictions of the stipulated toy process, not measurements of the repressilator. The example supplies a falsifiable distinction between a graph-level comparison and a kinetic comparison.

A stronger study would pair actual fluorescence time-series measurements with a separately documented computational implementation, then prespecify an observation map and matched perturbations. It would record sampling interval, reporter delay, measurement noise, culture and operating conditions, phase-assignment rules, controller timing, and test-set separation. Period estimates alone could be an inadequate target if differences in fluctuation statistics or post-perturbation recovery are relevant to the scientific question. A graph correspondence could survive a failed kinetic comparison; the failed comparison would restrict its warrant to topology. Conversely, matching sampled time series over a short observation window could coexist with different latent regulatory or computational mechanisms, as the underidentification analysis of Section 7 anticipates.

A meaningful negative control would use a controller or biological abstraction with a similar output period but a different permitted state transition or recovery mechanism. If the declared translation still classifies this control as equivalent for the task of predicting recovery, the proposed categories are insufficiently discriminating. This requires holding constant the evaluation target while altering the mechanism, rather than retrospectively choosing a convenient similarity metric. The interpretive remainder includes cellular metabolism, molecular regulation, implementation-specific hardware and code, and the distinct experimental interventions available in the two systems. Neither a common diagram nor equal oscillation periods settles those differences.

9.3 Ecological Food Webs and Resource-Flow Models

Ecological networks make visible another ambiguity of structural correspondence. Food-web diagrams represent feeding relations among selected ecological entities, yet graph edges can be assembled under varying sampling and taxonomic decisions. Dunne, Williams, and Martinez examined network properties across sixteen food webs and showed the importance of connectance and web size in interpreting their topology [8]. That empirical analysis motivates care in transferring claims based on degrees, path lengths, or other graph signatures. The present section makes no new estimate of food-web structure from their datasets.

A contrasting artificial system might be a resource-flow network in which material inputs are transformed into outputs by specified production rules. Both descriptions may be represented as directed networks. A proposed correspondence could relate ecological feeding connections to technological input dependencies. For the restricted task of comparing topological reachability, a carefully defined edge-to-edge mapping may preserve selected paths. The respective edges nevertheless encode different kinds of evidence and production constraints. An observed feeding relation need not specify the quantity, timing, or nutrient composition of the transfer; an engineering input specification may require precisely those quantities before a transformation is feasible.

The material comparison therefore needs an inventory or flux description alongside the directed graph. In ecological work this may involve biomass, energy or nutrient flow, feeding rates, turnover, and conditions of renewal. In an engineered resource process it may involve stoichiometric coefficients, finite stocks, waste flows, and constraints on simultaneous production. A hypothetical network with a single consumable resource enabling two alternative production branches illustrates the limitation of counting possible paths: each branch can be individually accessible while their simultaneous products remain infeasible. Section 5 establishes the relevant one-round resource-reservation criterion. The present case asks how an actual investigator would identify the material constraints needed to apply it, rather than restating the criterion as a new result.

Ecological comparison additionally raises questions of disturbance, persistence, and regeneration. Holling's influential discussion of resilience and stability identifies a historical research tradition concerned with the behavior of ecological systems under change [17]. The bibliographic reference motivates the distinction here; the present paper does not attribute any newly calculated resilience index or universal mechanism to that work. A claim that two ecological or technical networks have comparable resistance or recovery properties would require a specified disturbance, an admissible family of response trajectories, a definition of continued operation, and an observation horizon. The terms *resilience*, *efficiency*, and *stability* should retain their study-specific operational meanings.

An empirical design could select two well-characterized networks with similar connectivity statistics, then estimate relevant quantities under matched, physically feasible perturbations. The alternative explanations would include unobserved trophic interactions, missing resource stocks, distinct renewal rates, and inconsistent taxonomic or functional aggregation. The comparison should be repeated at more than one grouping resolution where data permit, since an apparent invariant may depend on the choice of species groups, process units, or boundary conditions. Preservation of a selected network statistic would count as a limited topological finding. Persistence of material fluxes and response capacities would require additional tests with quantified uncertainties. Possible ethical constraints on ecological interventions are themselves part of the admissible experimental design.

9.4 Grammatical Annotation and Artificial Parsing Systems

Linguistic comparison is mediated by representational decisions that can be unusually visible in the data. Universal Dependencies (UD) provides a community-developed framework for cross-linguistically consistent annotation of word segmentation, morphological properties, and syntactic dependency relations. The UD v2 resource description explicitly discusses the annotation scheme and its multilingual treebanks [23]. This supplies a documented example of partial standardization across languages rather than evidence that grammars are interchangeable or that the chosen labels exhaust the structure of natural language.

Consider a comparison between a human-annotated corpus in one language and the output of a computational dependency parser applied to another. At one level, the two representations can be mapped to a shared family of dependency labels. At another level, segmentation policies, morphological categories, construction frequencies, attachment conventions, and the sources of annotation disagreement may differ. A label-preserving correspondence can be useful for evaluating a selected annotation task; its validity depends on whether the label has sufficiently comparable operational content in the constructions under study. The same written label may cover different distributions of constructions, and a different label

need not by itself imply a different generative capacity.

The comparative task should therefore name its target. If the target is agreement with a held-out, consistently annotated evaluation corpus, relevant evidence may include per-construction error rates, inter-annotator agreement, sampling composition, and parser uncertainty. If the target is the generative organization of grammatical expressions, additional evidence from language use, structural alternations, or explicit grammatical hypotheses may be necessary. Annotation accuracy is then a measurement of performance on a particular representational convention, with its relationship to linguistic explanation remaining a substantive question.

An appropriate case design would fix versions of the annotation guidelines and record deviations and language-specific extensions, then compare the consequences of alternate segmentations or attachment conventions for the selected research task. The record would distinguish changes to the objects being annotated from changes to the underlying data, model, or linguistic hypothesis. A deliberately chosen negative control might swap a set of labels while leaving the tree shape unchanged. A topology-only comparison would classify the resulting graphs as structurally similar, while a task requiring interpretation of grammatical relations would detect the displaced distinctions. The failure is informative because it identifies exactly which aspect of the representation was required for the intended inference.

This example also exposes institutional dimensions of epistemic translation. Annotators, corpus maintainers, and language communities may have different opportunities to define categories and correct conventions. Their roles cannot be inferred from a treebank alone. The case study should document how corrections are proposed and resolved, especially when annotation resources differ sharply among languages. Section 8 provides the narrower account of negotiated translation relevant here; Section 10 will separately investigate institutional authority over shared comparative categories. The grammar case remains an empirical research proposal: no cross-language scoring or direct corpus analysis has been carried out for the present manuscript.

9.5 Provenance, Joint Feasibility, and Cross-Case Residuals

The preceding cases allow a common methodological comparison without assimilating their subject matters. A cyclic pattern in genetic regulation and computation, a directed relation in food webs and production systems, and a dependency graph in linguistic annotation are describable through graph-like structures. The respective relation types, state spaces, observation interfaces, and interventions nonetheless differ. The appropriate comparison may preserve a graph property while leaving other structures outside the translation domain. Documenting those remainders is part of a successful scientific result, because they delimit the claims supported by the chosen correspondence.

Historical provenance supplies a further discriminating factor. A current output might be produced by a biological regulatory network with one developmental and material history or by a controller with a different implementation and tuning history. Identical current food-web topology may coexist with differing disturbance histories and ecological stocks. Similar annotation graphs may arise through different annotation conventions or computational training procedures. In each setting, the empirical question is whether historical differences alter the target prediction or explanation. As Section 6 established under its formal assumptions, histories sharing a present description preserve selected continuations only when the relevant successor correspondence is maintained. The empirical task is to determine whether the observation state is sufficiently informative for that condition to be plausible.

The cases also reveal why joint generative feasibility requires domain-specific models. Multiple regulatory activities can compete for molecular resources; ecological processes share material stocks and affect their replenishment; a parser can share computational memory and processing resources among tasks. A common mathematical language for constraints may support comparison, yet the relevant units, conservation rules, and mechanisms must be supplied by specialist investigation. The researcher should avoid treating these examples as evidence that all resource relations instantiate one privileged ontology. An alternative interpretation may hold that the shared constraint language is instrumentally useful while its domain-specific meanings remain irreducible.

Cross-case residuals can be specified prospectively. A case record may identify unrepresented physical degrees of freedom, unmeasured historical variables, interpretive categories without stable equivalents, or institutional exclusions from the design of the comparison. If a candidate mapping performs satisfactorily for its declared task, its residuals remain visible as limitations of scope. If it fails, the investigator can ask whether the failure arose from a defective correspondence, an incompatible experimental regime, an inadequate measurement, or an erroneous expectation of preservation. These alternatives require distinct revisions and should therefore be recorded separately.

9.6 Discriminating Tests and Specialist Baselines

The evaluation of G4 requires comparisons against established approaches on identical evidential tasks. For the biological and computational case, state-space modeling, dynamical systems, and stochastic process analysis are the specialist baselines. For ecological material flows, network analysis, ecological modeling, and mass- or energy-balance descriptions are relevant. For linguistic annotation, corpus linguistics and parser evaluation already provide task-specific standards. The G4 framework should be considered successful only if its combined record clarifies a previously obscure warrant, reveals a transfer that specialist comparison would otherwise leave implicit, or improves the specification of a test. The present paper has yet to demonstrate any such comparative performance advantage.

A discriminating design begins with a declared question and at least two candidate explanations. Consider the regulator/controller comparison: one explanation attributes a recurring trace to a shared cyclic abstract structure, while another locates predictive differences in timing and perturbation response. The time-to-transition observations of Equation 22 distinguish constructed models with different rates even under identical support graphs. Actual biological and computational observations would require additional modeling and calibration. In the ecological case, a link-count hypothesis competes with an inventory-dependent feasibility hypothesis; measuring relevant fluxes and testing feasible simultaneous outcomes can discriminate them. In the linguistic case, an apparent shared relation label competes with a construction-dependent interpretation; testing alternate conventions on held-out cases can expose the difference.

Negative controls need to challenge the claimed preservation relation. Controls that simply produce large differences under every description add little information about the boundary of comparability. A more informative control matches a convenient surface feature while altering a property required for the target inference. Likewise, sensitivity analysis should change one declared mapping or sampling choice at a time where feasible, then examine which conclusions remain supported. If multiple changes are necessary, the evaluation should state the dependency between them rather than attributing failure to a single factor.

The distinction between predictive agreement and explanatory transfer is especially important. Two models can yield similar predicted observables over the tested horizon while assigning different mechanisms to hidden components. A task that requires only forecasting within that horizon might regard the models as adequate substitutes under measured conditions. A task concerning intervention, restoration, or material feasibility would require different evidence. A comparison is thus evaluated by its fit to a declared question and available tests, preserving the possibility of partial success without converting local adequacy into global equivalence.

9.7 Comparative Case Records and Replication Conditions

Table 8 condenses the proposed investigations into a structured case record. Each row distinguishes a selected candidate invariant from the additional evidence needed for a stronger inference. The table lists research requirements and failure probes; it reports no new measured outcomes. Its entries could guide a prospective comparative study or a retrospective audit of existing datasets, subject to data access and domain-specific ethical requirements.

A minimally reproducible record would retain the versioned data and coding conventions; the proposed mapping and its admissible domain; units, temporal and spatial resolutions, and resource assump-

Table 8: Prospective tests of selected cross-system correspondences; all outcomes remain to be evaluated.

Case and selected invariant	Additional discriminating evidence	A comparison-failure probe
Regulatory oscillator / controller: cyclic abstraction	Waiting times, fluctuations, perturbation response, calibration of phases	Matched graph and period with divergent recovery behavior
Food web / resource flow: selected connectivity	Fluxes, stocks, renewal constraints, disturbance responses	Similar connectivity with incompatible joint material feasibility
Treebank / parser: shared annotation labels	Segmentation, construction-specific agreement, alternative guidelines	Preserved tree shape with altered interpretation of relations
Historical comparison: matching current observations	Formation records, controlled changes, successor feasibility	Equal observed states with differing admissible continuations

tions; the declared evaluation target; the candidate models and negative controls; and the status of each source. When the source material involves sensitive human information, protected ecological sites, or proprietary systems, responsible access constraints may limit publication of raw data. A public specification of methods, synthetic examples, and appropriately aggregated findings can still support scrutiny, although they do not replace access to decisive evidence.

The evaluation should use predeclared success conditions. For example, a study may specify acceptable prediction error for a selected observable, a tolerance for conserved quantities, or the classes of grammatical relations whose interpretation must survive translation. Choosing these tolerances requires disciplinary judgment; they cannot be fixed universally by G4. Where uncertainty in the source or mapping is material, a comparison may yield an identified set of outcomes rather than a single verdict, consistent with Section 7. A result that remains unresolved is an admissible outcome of a disciplined comparison.

9.8 Interpretive Consequences of the Comparative Cases

The cross-domain cases support a limited but important methodological conclusion. Similarity of observed patterns can motivate a comparison, while stronger claims depend on preserved operations, compatible conditions, and evidence of the selected consequences. A directed cycle, a network statistic, or an annotation label can serve as a stable object of study within a declared description; each becomes insufficient when the research question reaches beyond what that description retains. This is a claim about the structure of comparative reasoning, illustrated by constructed cases and supported by distinct specialist literatures, rather than an empirical generalization about the systems themselves.

The examples also offer a way to evaluate the research programme critically. If a specialist baseline, given the same question and evidence, records the same preserved structure, the same loss of information, and the same revision conditions with equal clarity, G4 has not demonstrated added explanatory discrimination for that task. Its vocabulary may still be useful for organizing interdisciplinary inquiry, although such utility warrants separate evaluation. A more demanding test would compare independent teams using a G4-style comparative record against teams using established discipline-specific protocols, with preregistered evaluation criteria and reviewers competent in the relevant fields. This remains a proposal rather than a completed validation study.

Finally, comparative categories are chosen and maintained through research practices. A requirement to measure kinetic fluctuations, identify ecological resource boundaries, or distinguish grammatical relation types determines which structures become visible. These choices can have scientific and institu-

tional consequences, especially when some participants or affected communities have limited influence over what counts as relevant evidence. The next section therefore examines *authority over comparative categories*: how metric selection, data access, institutional standing, and opportunities for correction condition the legitimacy and reliability of cross-system comparisons. The present section has supplied bounded case specifications and testable limits while retaining the distinct historical and material character of each domain.

10 Authority over Comparative Categories

Comparative inquiry requires decisions about which entities enter a study, which differences count, how evidence is obtained, and which results are treated as relevant to a practical judgment. These decisions have a dual character. They are methodological conditions for meaningful inference and, in institutional settings, potential sources of authority over participants whose circumstances are represented. The objective of this section is to examine the relationship between comparative adequacy and the institutional conditions under which categories become authoritative. The discussion proceeds through classification design, commensuration, evidence access, standing, metric sensitivity, correction procedures, and bounded cross-domain examples. It develops a simple robustness proposition and a procedural evaluation framework while keeping mathematical validity, institutional effectiveness, and normative legitimacy analytically distinct.

10.1 Classification Decisions within Comparative Inquiry

A comparison begins with a selection, even where its categories have been inherited from a mature discipline. Investigators determine the entities or practices to be grouped, the descriptions under which they are recorded, the transformations to be observed, and the boundaries of the data. They may additionally specify equivalence relations, measurement instruments, acceptable uncertainty, and a rule for combining results. Each choice determines which differences will remain legible in the resulting comparative record. Section 3 treated a mapping as a mathematical object with a declared domain and preservation requirements. Its practical construction also depends upon who is empowered to select that domain and revise its definition.

The social life of classifications has received sustained attention. Bowker and Star examine the organization of practices through categories and standards that become embedded in information infrastructures [5]. Their analysis provides a useful historical and sociological counterpoint to a formal description of equivalence classes. A quotient of a mathematical structure is specified through an explicit relation; institutional classifications often acquire their meanings through histories of use, negotiated conventions, technical constraints, and administrative routines. A category can be formally coherent for its declared purpose while having consequences that extend beyond that purpose. Conversely, a controversial category can remain indispensable for coordination within a particular limited task. The analytical question concerns the category's warranted scope and the institutional pathways through which that scope is established.

For a comparative episode, one may record a provisional *category specification* as a tuple $\mathcal{K} = (\Omega, q, f, \mu, a, \mathcal{E}, \Gamma)$. Here Ω is the declared domain of cases, q the guiding question, f the proposed coding or translation map, μ the observation procedure, a an aggregation or decision rule where one is used, $\mathcal{E}$ the evidence-access conditions, and Γ the procedure for proposing, examining, and recording revisions. This notation serves as an audit checklist. It is neither a universal ontology of categories nor a causal model of their institutional effects. In particular, the specification of Γ documents a procedure; it offers no immediate guarantee that a less powerful participant can exercise it effectively.

A category specification should be read in conjunction with its provenance. A classification inherited from one research programme may travel across jurisdictions, languages, ecological settings, or technical platforms. Its terminology can remain stable while the activities associated with its codes change. The resulting mismatch may be an error of measurement, a change of operational regime, a difference in

interpretation, or a substantive difference between systems. Those possibilities call for distinct inquiries. Recording the origin of category choices helps prevent an apparently neutral numerical output from concealing the historical decisions that made it possible.

10.2 Commensuration and the Choice of Metrics

A shared metric can facilitate comparison by translating different observations into a common form. Its analytical value depends upon the intended question and the properties preserved during translation. Espeland and Stevens describe commensuration as a social process that renders different entities comparable by means of a common metric, with cognitive and political consequences [10]. Porter, examining the historical authority of quantification, places numerical objectivity within institutional and social practices rather than treating the public standing of numbers as self-explanatory [26]. These accounts help identify an additional research object for G4: the production and circulation of the *authority to declare two heterogeneous cases commensurable*.

Consider two research infrastructures assessed through an index aggregating accuracy, energy use, access costs, and recovery capacity. Each indicator can be operationalized, provided its observation conditions and units are specified. Yet the move from a vector of separate measurements to one scalar ranking requires a further rule. Weighting, normalization, and the treatment of missing information can alter the result. A high score can therefore be evidentially meaningful for a specified purpose while remaining a poor basis for a claim about a system's general worth. For the present inquiry, the scalar index is an instrument for a selected comparison, with its legitimate uses requiring separate argument.

In some domains, comparison can be conducted without scalar commensuration. Pareto-type partial comparisons, explicit sets of permissible weights, interval-valued evidence, or a collection of separately reported criteria may retain distinctions that a single ranking suppresses. These alternatives impose different burdens upon readers and institutions. An interval may be more faithful to unresolved uncertainty yet less convenient for administrative allocation. A partial order can display incomparability yet require additional public reasoning for a concrete decision. A single index can aid coordination while obscuring variation in what its components mean. Methodological adequacy involves examining these trade-offs in relation to the actual use of the comparison, rather than treating one representational format as intrinsically superior.

The risk of assigning external authority to a chosen metric is especially salient when indicators circulate beyond the research context for which they were designed. Merry's analysis of human-rights and governance indicators offers documented examples of how quantified categories carry institutional theories and reshape the interpretation of complex social circumstances [22]. G4 draws a narrower methodological lesson: the evidence supporting a ranking and the evidence supporting its practical deployment are different evidential objects. For decisions affecting persons, populations, or ecosystems, the basis of inclusion, the foreseeable consequences of omission, and the conditions for reconsidering the metric deserve explicit evaluation.

10.3 Evidence Access and Asymmetric Observability

Authority over comparison can arise before a metric is applied. A participant or institution with privileged access to records, instruments, computing infrastructure, or interpretive expertise may determine which comparative models are admissible. Another participant may possess relevant local knowledge but lack the means to introduce it into the evaluation. Such asymmetries affect both the inferential accuracy of a comparison and the distribution of opportunities to challenge it. They can also be reproduced by apparently technical decisions about documentation formats, data retention, or model interfaces.

Section 7 showed that a comparative target is identified only when it is constant across the models compatible with an evidential record. That result assumes the evidence map and the candidate family have been specified. Institutional control over these elements therefore conditions which uncertainty is visible. A model excluded because its data were never collected differs epistemically from a model excluded because its predictions failed a discriminating test. The former raises questions about the construc-

tion of the admissible evidence set; the latter concerns performance relative to an established procedure. A responsible audit records the distinction.

Several forms of evidential asymmetry should be separated. *Acquisition asymmetry* concerns who can generate records; *interpretive asymmetry* concerns who can decode or contextualize them; *admission asymmetry* concerns whose records enter a formal evaluation; and *revision asymmetry* concerns who can trigger reconsideration of a settled category. These distinctions can overlap, but their remedies differ. Open data may improve acquisition and interpretation under some conditions, while sensitive records require lawful limitations. Translation assistance may improve access without resolving the authority to decide which interpretations count. An independent review can evaluate errors while leaving the underlying ontology of the comparison contested.

The presence of unequal access does not itself establish that a comparison is false. An instrument may be reliable even where its operation is restricted, and specialized interpretation can be essential for scientific validity. The appropriate inquiry asks which claims require independent replication, which access restrictions are substantively justified, and whether an affected party has a credible pathway to introduce contrary evidence or obtain review. Epistemic accountability may therefore require a combination of methods: calibration records, provenance documentation, controlled disclosure, external replication, domain-specific expertise, and meaningful mechanisms for correction.

10.4 Standing, Representation, and the Limits of Agreement

A comparative category can be widely accepted because it solves a coordination problem, because a disciplinary community finds it useful, or because an institution mandates its use. Those routes confer different forms of authority. Scientific endorsement concerns the adequacy of a method within a relevant field; administrative authorization concerns a rule of institutional practice; political or ethical legitimacy concerns the justified exercise of power over affected participants. Agreement in one domain supplies limited evidence for the others. Section 8 distinguished interpretive agreement from adequate reciprocal translation; the same discipline applies to institutional ratification.

Longino's account of socially situated inquiry is useful for understanding why critical interaction contributes to the assessment of evidence, while Fricker's analysis of epistemic injustice supplies vocabulary for considering patterned failures to credit or interpret some human participants' contributions [19, 12]. G4 uses these contributions as philosophical resources rather than as proof that a single procedural design works across all institutional settings. An affected participant may have relevant knowledge about the consequences of a category yet lack technical expertise in its measurement model. A specialist may be well placed to evaluate measurement validity yet unfamiliar with the circumstances in which a score will be deployed. A legitimate comparative practice may need to accommodate both forms of contribution without conflating their domains of competence.

Representation becomes especially difficult when the compared systems include animals, ecosystems, or artificial infrastructures. Research institutions often rely upon human advocates, domain experts, and legal or regulatory procedures to represent affected interests. Such representation should be evaluated according to its authorized scope and evidential adequacy. Its existence does not establish that the represented entity endorses the categories or possesses a particular moral or epistemic status. Likewise, the ability of an AI system to generate fluent objections provides an observable communicative event, but its normative significance depends on further assumptions and institutional context. The present analysis maintains ontological suspension regarding contested forms of subjecthood and agency.

An evaluative record can distinguish the initiator of a comparison, its technical designer, the custodian of its data, the authority applying its results, and the people or systems whose conditions are affected. These roles may be occupied by one organization or distributed across several institutions. Their separation improves the diagnosis of a disputed comparison: an erroneous encoding may require technical repair, an inadequate authorization may require governance review, and an inappropriate use beyond the declared domain may require limiting or retracting the application. No single procedure resolves these different kinds of failure by definition.

10.5 Sensitivity of Rankings under Category Revision

The philosophical significance of metric selection can be made more precise through a limited quantitative exercise. Consider two systems a and b , each represented by a profile $z_a, z_b \in [0, 1]^d$ of *declared* measurements or scores, with the orientation of every coordinate fixed so that larger values indicate better performance on that particular criterion. A nonnegative weight vector w with unit sum defines a task-specific score $s_w(i) = \sum_{k=1}^d w_k z_{ik}$. This is a formal model of scalar aggregation; it does not determine whether the profiles are comparable, which criteria are relevant, or whether the weights are legitimate.

Proposition 10.1 (Margin stability under bounded metric revision). *Let $z_a, z_b, z'_a, z'_b \in [0, 1]^d$ and let w, w' be nonnegative vectors with coordinates summing to one. Suppose $\|z'_i - z_i\|_\infty \leq \delta$ for each $i \in \{a, b\}$ and $\|w' - w\|_1 \leq \varepsilon$. Write $\Delta_w = s_w(a) - s_w(b)$ and $\Delta'_{w'} = \sum_k w'_k (z'_{ak} - z'_{bk})$. Then Equation 23 holds. Consequently, a strictly positive original margin greater than $\varepsilon + 2\delta$ remains positive after the declared revisions.*

The bound referred to in Proposition 10.1 is displayed in Equation 23.

$$|\Delta'_{w'} - \Delta_w| \leq \varepsilon + 2\delta. \quad (23)$$

Proof. Add and subtract $\sum_k w'_k (z_{ak} - z_{bk})$. Since every coordinate difference $z_{ak} - z_{bk}$ belongs to $[-1, 1]$, the term arising from a change in weights has absolute value at most $\|w' - w\|_1 \leq \varepsilon$. The two profile-perturbation terms have total absolute value at most $\sum_k w'_k (\delta + \delta) = 2\delta$. The triangle inequality yields Equation 23. If $\Delta_w > \varepsilon + 2\delta$, the perturbed difference remains positive. $\square$

This elementary inequality is closely related to standard robustness reasoning in multi-criteria analysis. Its contribution to G4 is diagnostic: an asserted comparative ranking can be accompanied by a declared range of measurement and weighting revisions. The numerical margin states a conditional stability result within one model of scalar aggregation. It establishes neither a universal ranking nor a justification for the chosen dimensions. A procedure that changes the meaning of a criterion or introduces a new excluded group falls outside the proposition's fixed-coordinate assumptions. The inequality also assumes fixed coordinate alignment and the same normalization interval for both profiles; it offers no protection against category substitution or a change in what a score is intended to measure.

A complementary observation concerns incomparability. If one system performs better on one coordinate and worse on another, a weight concentrated on the first coordinate and a weight concentrated on the second can reverse the ranking. When one profile is coordinatewise at least as high as another, every nonnegative weighted sum preserves the corresponding weak order. These are familiar properties of weighted aggregation. They explain why the presence or absence of a robust ranking must be interpreted together with the admissible weight set and the provenance of the underlying profiles. For a comparison with uncertain measurements, it may be preferable to report intervals and unresolved cases rather than produce a seemingly definite ordering from one arbitrary set of weights.

10.6 Contestability and Correction of Comparative Claims

A comparison becomes accountable when its intended beneficiaries and affected parties can identify the claim being made, the evidence supporting it, and the procedure through which errors or inappropriate applications may be challenged. Contestability has both epistemic and material conditions. Publishing a formula can reveal how an index is calculated while leaving its input data inaccessible; offering a complaint channel can appear inclusive while giving its reviewers limited authority to change an outcome. A useful review therefore examines both what the procedure permits formally and what participants can realistically do under their circumstances.

A practical correction sequence can distinguish four stages. An initial challenge specifies the category, decision, or datum in dispute. Evidential review identifies whether the dispute concerns an inaccurate measurement, an inadmissible translation, a missing case, an unsuitable aggregation, or an inappropriate practical use. The responsible body then determines an available response within its legal and

institutional competence. A final record explains the decision and the criteria for subsequent reconsideration. These stages serve as a provisional analytic framework rather than a claim that administrative systems universally follow them.

Corrective actions may differ significantly. A mislabeled case can be recoded while leaving the underlying metric intact. A comparison that omitted a relevant variable may require recalibration or renewed observation. A scalar score that conceals an ethically significant difference may need to be supplemented with disaggregated results or withdrawn from the decision for which it was used. Where an entire category was selected without appropriate participation, redesign may require a more extensive institutional procedure. In each case, correction should preserve a versioned record so that a revised category can be related to the earlier claims and any decisions made under them.

There is a further distinction between *reviewability* and *effective remedy*. A challenge may successfully establish an epistemic defect without yielding authority to reverse an institutional decision or repair its material effects. Conversely, an institution may provide a pragmatic accommodation without settling the scientific disagreement. The relevant mechanisms, legal duties, and burdens of proof vary by context. G4 can clarify the type of defect and the evidence needed for correction while leaving context-specific legal and ethical judgments to the appropriate institutions and affected participants.

10.7 Comparative Episodes in Ecological and Artificial Systems

The methodological concerns become more concrete when categories travel between scientific and institutional settings. Consider a hypothetical coastal-monitoring programme that compares two habitats using a common “ecological recovery” score. One site has rapid return of a selected indicator species, while another preserves a slower but more diverse pattern of regeneration. A single score constructed from species counts, biomass, and disturbance frequency could rank the sites differently depending on weighting and observation intervals. That variation does not establish a defect by itself: different scores may answer different ecological questions. The comparative record should declare the target of recovery, its timescale assumptions, affected ecological functions, and the reasons for the selected indicators. It should also identify alternative interpretations that could change a management recommendation.

The related literature on resilience and food-web structure, discussed in Section 9, supplies disciplinary examples of why ecological function and network topology deserve distinct measures [17, 8]. The present coastal-monitoring episode remains hypothetical, with no reported field measurements. Its value is to expose the places at which authority can enter: the selection of monitored sites, the determination of baselines, the choice of a management objective, and the eligibility of different forms of ecological knowledge for review. A responsible research design would seek specialist ecological validation before translating any comparative index into intervention.

A second hypothetical episode involves two AI research infrastructures assessed for accessibility and research productivity. A comparison based on publication counts and benchmark throughput may favor one infrastructure, whereas a comparison including operating cost, language coverage, and the capacity of researchers to export and reproduce their analyses may favor the other. These are genuinely different evaluative objects. The challenge arises when an institution treats a score designed for throughput as an authoritative statement about overall scientific opportunity. A researcher whose work depends on an underrepresented language or a specialized data regime may have relevant evidence that the current metric omits. A category audit would examine how such evidence can be introduced and whether a revised instrument changes the comparative claim.

A third episode concerns multilingual annotation programmes. A shared dependency label facilitates certain structural comparisons across languages, as the Universal Dependencies programme illustrates at the level of curated representational practice [23]. Yet administrative or research uses of a single accuracy metric may privilege annotation conventions that are more richly represented in the training material. A responsible evaluation should distinguish the technical agreement of an annotation system from interpretive adequacy for individual linguistic traditions. Proposals for correction may include language-specific error analysis, independent annotation, or the preservation of multiple warranted analyses. The present

Table 9: Authority points and evidential records for revisable comparative categories.

Authority point	Record to preserve	Question for review
Category boundary	Included and excluded cases; revision history	Whose relevant differences remain visible?
Observation and translation	Instruments, coding maps, calibration, uncertainties	Which claims survive the selected representation?
Aggregation and metrics	Scales, weights, omitted dimensions, sensitivity results	Which rankings are stable within justified alternatives?
Evidence custody	Access rules, provenance and reproducibility limits	Who can introduce or test contrary evidence?
Institutional application	Decision scope, responsible authority, affected interests	Does the use remain within its evidential and legal warrant?
Challenge and correction	Review access, responses, new categories, versioned outcomes	Can identified defects be examined and remedied?

discussion specifies a research design, not measured differences among languages or parsers.

Across these three episodes, the relevant category authority takes different institutional forms. Ecological monitoring involves scientific expertise and possible public stewardship decisions; AI infrastructure assessment involves technical operators and research institutions; multilingual annotation involves collaborative standards and linguistic interpretation. The comparative framework remains useful precisely when it avoids forcing these arrangements into a single governance template. Its practical role is to make category decisions, evidential limits, and correction opportunities visible to the inquiry in which they arise.

10.8 Revisable Comparative Standards

A comparative standard can remain scientifically productive while undergoing revision. Calibration changes, new evidence, a changed target population, or a newly recognized interpretive distinction can require an update. The intellectual task is to preserve the lineage of the comparison: which earlier results remain supported, which become conditional, and which are withdrawn or reinterpreted. A change in category boundaries is therefore more than a technical patch when it alters the membership of the compared groups or the consequences attached to their rankings. It may also change the social and material distribution of the capacities needed to contest future assessments.

Table 9 sets out a provisional record for evaluating such revisions. The rows identify different sites of authority, including observational selection, commensuration, evidential access, decision use, and correction. The suggested records are methodological aids for designing a comparative inquiry, not a validated institutional accreditation scheme or a universal ethical checklist.

The section's conclusion is methodological and qualified. A comparison may be mathematically exact within selected categories, empirically useful for a declared predictive task, and institutionally contested in its application. Each description refers to a different evaluative question. Category authority becomes relevant to G4 because the production of comparability is itself a situated practice that distributes visibility and opportunities for correction. Making that practice inspectable can improve the accountability of comparative claims; demonstrating improvements in actual institutional outcomes requires further empirical study. Section 10 thus supplies the institutional complement to the formal correspondence and reconstruction analyses. The following section develops research limitations and evaluation protocols capable of testing their combined adequacy.

11 Research Limitations and Evaluation Protocols

A comparative framework gains scholarly significance through the distinctions it enables researchers to justify, revise, and test. The present section evaluates the reach of G4's account of cross-relational

invariance and specifies research procedures capable of challenging it. Its objective is to distinguish established methods from the proposed coordination of those methods, identify principal sources of inferential failure, and define evidence that would support either a narrower or a stronger version of the programme. The discussion proceeds through specialist alternatives, formal provenance, measurement and translation, causal interpretation, case selection, comparative performance, institutional accountability, and explicit revision criteria. Formal notation is used only to clarify the evaluation object. The proposed protocols are methodological designs; no experiment testing the overall G4 framework has yet been conducted.

11.1 Explanatory Contribution and Specialist Alternatives

The central methodological risk is redescribing established techniques through a new vocabulary without improving an actual investigation. Graph theory already evaluates reachability and structural correspondence; bisimulation and process semantics address selected forms of behavioural equivalence; resource-sensitive formalisms distinguish alternative accessibility from joint feasibility; latent-model identification and inverse-problem analysis examine observational ambiguity. In particular, the techniques invoked in Sections 3, 7, and 9 inherit substantial capabilities from these disciplinary resources [28, 4, 21, 18]. A comparison that is completely and adequately handled by one specialist model should be credited to that model.

G4’s distinctive proposal therefore lies at the level of *comparative warrant coordination*: making explicit which kind of correspondence is asserted, which information supports it, which remains omitted, and how a conclusion changes when the research task changes. This claim concerns the organization and criticism of inquiry. It remains separate from claims of superior prediction, algorithmic efficiency, or discovery of mathematical invariants. Such advantages would require independent demonstrations against relevant baselines. A technically elaborate audit might even make research less effective when the additional record does not change any inferential decision or when its reporting burden exceeds its value.

Three rival explanations are particularly important. First, a specialist investigator may already make the required distinctions implicitly, making G4’s record an alternative exposition of existing good practice. Second, any observed improvement may result from additional time, calibration, or expert consultation rather than the proposed comparative categories. Third, a richer framework may improve written accounts of uncertainty while failing to improve predictions, replication, or warranted decisions. These possibilities suggest an evaluation design in which procedural effort, domain knowledge, and information access are recorded and, where feasible, matched across conditions. A favorable comparison with a deliberately weak baseline would provide little evidence for the programme’s added value.

A research claim should specify the level of comparison at which improvement is sought. *Explanatory discrimination* concerns the correct separation of rival mechanisms. *Predictive discrimination* concerns held-out observations under declared conditions. *Interpretive adequacy* concerns whether the mapping preserves meanings or distinctions relevant to its task. *Institutional accountability* concerns documented opportunities to challenge and correct the use of comparative categories. Each requires its own evidence; combining them into a single achievement score would conceal important differences among their aims.

11.2 Formal Provenance and the Scope of the Results

An audit of the formal propositions must distinguish their mathematical validity from their provenance and the scientific inferences drawn from them. The manuscript contains fourteen numbered propositions across Sections 3–10. Their proof obligations are local: each concerns declared sets, maps, relations, resources, observations, or bounded perturbations. Where the proposition is an elementary characterization, its proof establishes that characterization within the stated model; it does not establish a newly discovered mathematical structure. The accompanying verification scripts examine selected finite instances and counterexamples. Independent proof review remains a separate scholarly requirement.

Table 10: Formal provenance and remaining warrant boundaries for the manuscript’s propositions.

Formal domain	Established mathematical baseline	Further comparison requiring evidence
Local relational maps (Section 3)	Isomorphisms, saturated quotient relations, composition of partial homomorphisms	Mapping semantics, selected domains, closure and regime alignment
Preserved quantities and rankings (Sections 4, 10)	Intertwining, pullback of invariants, deterministic perturbation inequalities	Calibrated measurements, common units, justified weights and interpretive scope
Resources and histories (Sections 5, 6)	Independent-product reachability and labelled successor congruence	Joint material feasibility, archived events, and predictive or causal adequacy
Inference and translation (Sections 7, 8)	Factorization on observational fibres, set inclusion, restricted quotient injection	Candidate-family completeness, valid probes, effective reciprocity and category authority

Table 10 organizes these propositions by their mathematical starting point and the inferential limit that remains after proof. This classification prevents a graph-theoretic correspondence, for example, from being reported as equivalent biological dynamics, and prevents point identification within a selected model family from being reported as identification across every plausible model. The specialist literature on behavioural equivalence, open Petri nets, and statistical identification provides important antecedents [28, 4, 21, 2].

The central methodological distinction is therefore between *formal implication*, *model adequacy*, and *practical warrant*. A correct result about a selected relation licenses a conclusion about that mathematical relation. Using it to compare experimental, ecological, linguistic, or institutional systems requires separately supported interpretations of the variables, admissible operations, and observed conditions. A well-documented correspondence may be valuable even where the relevant wider interpretation remains unresolved. The claim register accompanying the editable source records, for each proposition, its assumptions, an independent proof outline, the corresponding baseline, and remaining questions for external review.

11.3 Construct Validity across Heterogeneous Descriptions

A comparison can fail before any formal mapping is evaluated, because the observable quantities selected in its two domains may instantiate different concepts. Labels such as *resilience*, *capacity*, *complexity*, *productivity*, or *adaptation* acquire operational content through discipline-specific instruments and decisions. The presence of a common label supplies a hypothesis of comparability. Its empirical standing depends on a traceable connection between the intended concept, the particular operational measure, and the available evidence. Adcock and Collier’s account of measurement validity provides a relevant methodological precedent for distinguishing conceptual disputes from disagreements over measurement and for attending to contextual specificity [1].

The problem becomes acute across biological, ecological, computational, and linguistic settings. In the constructed comparisons of Section 9, a cycle may mean a recurrent regulatory pattern, a permitted controller transition, or an aggregate graph property. A count of cycles may be measured reliably in each description while supporting only a narrow cross-domain claim. Similarly, a shared dependency label in two linguistic corpora may be consistently annotated yet differ in how it relates to the languages’ grammatical traditions. Domain experts must evaluate the link between concepts and records before treating an invariant calculated from the records as an invariant of the underlying target.

A validation design should distinguish at least four layers. *Conceptual alignment* documents the task-relevant interpretation of terms in both systems. *Operational alignment* records the instruments and coding procedures used to produce comparable observations. *Transformation adequacy* tests whether a proposed correspondence preserves an explicitly chosen relation or quantity. *Inferential adequacy*

evaluates the conclusion drawn from the preserved property. An exact graph isomorphism may satisfy the third layer for a topology task while leaving the fourth layer unsupported for a claim about physiological recovery or institutional legitimacy. Section 10 adds the separate question of who can revise the adopted concept or operational boundary.

The appropriate negative controls are conceptually informative. One may construct or select cases with the same measured output and different known mechanisms, or cases with different surface descriptions and a well-supported common operational mechanism. A method that treats the first pair as globally equivalent or rejects the second solely because of a changed label demonstrates sensitivity to irrelevant features. The candidate controls need independent justification, however: calling two mechanisms *known to differ* already depends on source-level evidence and the chosen scientific account. A valid assessment should disclose those sources and the scope of the distinction they support.

A further limit concerns equivalence across observation resolutions. Increasing sampling frequency, adding a new channel, or revising an annotation scheme may divide a previously unified class of configurations. Comparative findings should therefore carry a versioned description of the observation map. Whenever the selected resolution changes, the analysis should re-evaluate its invariant claims rather than silently carrying them forward. Such discipline is especially important when latent information is unevenly available across compared systems.

11.4 Causal Identification and Intervention Design

Correspondence between observed patterns supplies a starting point for comparative explanation. Causal interpretation requires further assumptions about interventions, dependencies, background conditions, and the origin of observational differences. Distinct mechanisms can reproduce the same joint distribution of selected observables, and a successful translation of that distribution may leave the mechanisms underdetermined. The partial-identification analysis in Section 7 formalizes this problem at the level of candidate models. Its resolution in empirical practice depends on the availability and credibility of distinguishing evidence.

An intervention-based comparison should therefore state its target operation separately in each system. Perturbing a gene-regulatory network, interrupting a software controller, removing a link in an ecological food web, and modifying a grammatical annotation procedure are interventions of different kinds. Their comparison requires a defensible account of what is held fixed, which mechanisms are activated, and why any observed difference is attributed to the selected intervention. The causal-inference literature provides established tools for making such assumptions visible; Hernán and Robins, for example, organize inference around explicit causal questions and the conditions under which observational data can answer them [15]. Their framework does not, by itself, equate interventions across heterogeneous domains.

Three common confounding mechanisms deserve scrutiny. A *shared-input mechanism* can produce aligned outputs when two systems respond to one background variation. A *selection mechanism* can create similarity when only comparable successful cases enter the record. A *representation mechanism* can manufacture apparent agreement when different observations are standardized through the same coding rule. Evaluating these alternatives may require within-system perturbations, longitudinal observations, negative-control variables, or separate instruments with different measurement errors. Each design should state the intervention's feasibility and ethical constraints before it is interpreted as a causal test.

Causal evidence can also reveal important limits of invariance. A mapping that preserves equilibrium frequencies might fail to preserve responses to perturbation; a mapping that preserves reachability might fail to preserve intervention costs or transition rates. The appropriate revision is to restrict the target relation to what the evidence actually preserves. Such a restriction need not count as failure of the entire programme. It can constitute a successful discovery of the boundary of a specific correspondence. The stronger programme-wide failure would be persistent inability to articulate or test those boundaries in a way that improves upon existing specialist inquiry.

11.5 Historical Sampling and the Scope of Generalization

A cross-system comparison also depends on how the participating systems were chosen. Cases selected for striking surface resemblance tend to overrepresent successful alignments and underrepresent the situations in which translation fails. Historical dependence further complicates selection, because presently similar states may have arisen through different paths and may retain different feasible continuations. Section 6 supplies a formal sufficiency condition for preserving selected future possibilities; applying it requires observations or defensible models of the relevant histories.

Seawright and Gerring distinguish several strategies for purposeful case selection, including typical, diverse, extreme, deviant, and most-similar or most-different cases [29]. For the present inquiry, a *diverse-case sample* can expose whether a proposed correspondence relies on one domain’s conventions; a *deviant-case sample* can reveal the mechanism of an unexpected failure; a *most-similar comparison* may help isolate a candidate difference in resources or history. These are different research purposes. A handful of theoretically chosen examples cannot establish the frequency with which a proposed invariant appears across an undefined population of possible systems.

Historical records also have heterogeneous reliability. An ecological reconstruction can depend on incomplete species surveys and seasonal sampling; a technical platform’s record may omit earlier configuration changes; a linguistic corpus records only the texts, annotation conventions, and sociohistorical traces available to its compilers. A comparison should specify whether history is directly observed, reconstructed from proxies, or represented only by a set of compatible trajectories. A proposed common history-dependent capability becomes more tentative when its essential inputs are inferred from the same output that the comparison seeks to explain.

Generalization should proceed in stages. A within-case result may demonstrate that one correspondence works under a specified regime. A matched comparison can examine whether it survives a known change in representation or material constraint. A purposively varied set can expose different failures. A broader sampling programme would be needed to estimate prevalence, with a defensible sampling frame and an account of dependence among cases. The investigation should report which level its evidence reaches, including cases excluded by inaccessible data, incompatible instruments, or unresolved translation.

11.6 Comparative Performance on Held-Out Tasks

The framework’s practical research value can be examined with a protocol that compares its use against a credible domain-specialist workflow. Participants or research teams could be assigned the same collection of constrained comparative tasks and access to the same relevant materials. One condition would use the ordinary specialist procedures judged appropriate for each task. A second would additionally use a G4-style record of comparison scope, candidate mapping, preservation target, omitted conditions, rival models, and anticipated failure observations. Both workflows should receive adequate time for familiarization. Where random assignment is feasible and ethical, task order and analyst assignment could be counterbalanced; otherwise the design must disclose the principal sources of selection bias.

A task should present an observable judgment that can be independently assessed. For example, an analyst might determine whether two transition descriptions genuinely admit a jointly feasible production outcome, whether an observed equality persists under a prescribed intervention, or whether an annotation correspondence generalizes to held-out constructions. The task designer must establish a reference standard by proof, independent measurements, or a justified expert adjudication procedure. A reference standard that merely repeats the G4 classification would prejudge the comparison. Some conceptual and normative questions may lack a single correct binary answer; in those cases a suitable assessment should record reasoned alternatives and the grounds on which they are judged.

For a predeclared set of tasks $\mathcal{T}$, the first reporting unit should be the task-specific paired loss difference $d_t = \ell_t(a_t^S) - \ell_t(a_t^G)$, with the sign, scoring rules, and units of each ℓ_t documented. Comparisons across tasks require an additional commensuration decision. Raw differences measured in distinct units can be aggregated only after a justified normalization and a disclosed choice of task weights. Equation 24

gives one explicitly conditional descriptive aggregate, using positive scale constants s_t and nonnegative weights ω_t fixed before observing the comparative results.

$$\hat{\Delta}_{\mathcal{T}}(\omega, s) = \sum_{t \in \mathcal{T}} \omega_t \frac{\ell_t(a_t^S) - \ell_t(a_t^G)}{s_t}, \quad \sum_{t \in \mathcal{T}} \omega_t = 1. \quad (24)$$

The constants s_t may express a justified reference loss, a predeclared performance range, or another task-specific scale with a substantive interpretation; their selection requires an account of why one normalized unit is comparable to another. Equal weighting is one possible convention, rather than a neutral default. A positive value of Equation 24 indicates lower weighted, normalized loss for the G4-assisted workflow under the selected conventions. It cannot by itself establish superiority across heterogeneous epistemic objectives.

The danger of undocumented scales is visible in a two-task constructed example. Let the specialist-minus-G4 loss differences be $d_1 = 2$ and $d_2 = -1$. An unnormalized equal-weight average is 0.5, whereas measuring the first task in a unit ten times larger changes its numerical difference to 0.2 and reverses the average to -0.4 without changing either workflow's performance. This is a change of measurement units, not an evidential reversal. The same sensitivity can arise through a defensible disagreement over the importance of prediction, explanatory discrimination, or interpretive adequacy. Reporting d_t , s_t , and ω_t separately, with sensitivity analyses over justified alternative weights and scales, protects the analysis from an apparent gain created by arbitrary arithmetic.

The aggregate is a descriptive statistic. Its inferential interpretation depends on task selection, repeated or clustered observations, scoring reliability, assignment, and uncertainty estimates. A stronger protocol would report confidence intervals or other appropriate uncertainty assessments within the declared research design, and would retain disaggregated performance by domain. The two workflows should also be compared on time, data requirements, expert involvement, unjustified confidence, and the practical cost of maintaining the comparative record. These resource and accountability measures may be incomparable on a single scale; reporting them in parallel can be more informative than a forced total score.

A serious test needs held-out questions and difficult negatives. The analyst should encounter cases where a highly visible graph match coexists with incompatible kinetics, where two production paths compete for one consumable input, where distinct histories share a present observation, and where a seemingly stable ranking changes under a defensible metric revision. These examples can be designed from independently established formal models before the analysts see them. The evaluation should also include ordinary cases adequately resolved by specialist methods, so that the additional comparative apparatus receives no credit merely for reproducing a conventional result. Reporting calibration of analysts' uncertainty would be useful where confidence judgments are elicited in a consistent format.

The present manuscript has *not* conducted this comparative-performance evaluation. The finite verification scripts accompanying earlier sections check the mathematical consistency of selected propositions and counterexamples, rather than the performance of the G4 procedure in scientific practice. An assessment of practical improvement must be conducted separately, ideally with protocols available for external criticism before results are collected.

11.7 Epistemic Participation and Institutional Consequences

The evaluation of comparison cannot be confined to technical correctness when classifications affect the people or institutions they describe. Section 10 distinguishes the accuracy of a category from the authority to impose it and the legitimacy of its application. A methodologically successful metric can still be used outside the target domain that justified it. Conversely, a contested institutional rule may use a scientifically defensible classification for one narrow purpose. An evaluation protocol should report these situations separately rather than interpret disagreement as measurement error.

Practical scrutiny begins with evidence custody and review. Which records can be inspected independently? What lawful or ethical limits constrain disclosure? Which parties have standing to correct

an omitted context or challenge a translation? Are corrections versioned, communicated, and traceable to the conclusions they alter? These questions are informed by the work on classification infrastructures, indicators, and epistemic exclusion discussed in Sections 8 and 10 [5, 22, 12]. Their inclusion does not imply that one procedural template is suitable for every scientific field or institutional setting.

A comparative intervention can impose costs on participants whose ability to respond is uneven. Adding an experimental ecological disturbance may be scientifically discriminating while raising animal-welfare and conservation concerns. Releasing infrastructure records can aid independent evaluation while exposing confidential information. Asking linguistic communities to correct an imposed category may require time and expertise that existing projects have failed to resource. The research design should acknowledge these constraints before proposing a test. In some cases observational alternatives, simulations, controlled access, or participatory review can provide a more appropriate means of inquiry.

Institutional evaluation should include both the documentation of challenge procedures and evidence that they work in practice. An appeal route without accessible records, interpreters, or timely decisions may be only formally available. Meaningful assessment would examine whether challenges can change categories, whether corrections propagate into consequential decisions, and whether affected participants can obtain explanations. These outcomes require setting-specific empirical study. G4 provides an organizational rationale for recording them; it has yet to demonstrate that the rationale improves institutional accountability.

11.8 A Staged Protocol for Comparative Inquiry

A provisional protocol may help connect the preceding limitations to replicable research. It should include an explicit candidate-family sensitivity stage, testing whether identified comparisons survive the admission of plausible alternative mechanisms and mappings as described in Section 7.6. Any cross-task score should be accompanied by its normalized scales, task weights, and disaggregated results. Its aim is to expose the inferential chain before conclusions are announced, while allowing discipline-specific methods to determine the proper instruments. For each investigation, researchers should preserve a dated and versioned record of the comparison question, target and descriptions, relevant conditioning regimes, proposed transformations, evidence, alternative mechanisms, and the scope of inference. The record should identify the responsible researchers or institutions and the means of correcting material errors. It is a research aid, not an accreditation instrument.

Definition 11.1 (Task-Relative Comparative Warrant Record). A *task-relative comparative warrant record* is a versioned tuple $W = (q, \mathcal{M}, \Gamma, \tau, E, \mathcal{A}, \mathcal{V})$, where q specifies a comparison question; $\mathcal{M}$ identifies candidate systems and the selected descriptions; Γ declares admissibility conditions and relevant histories; τ specifies proposed correspondences and preservation targets; E identifies observations and their provenance; $\mathcal{A}$ records serious alternative accounts and anticipated discriminating evidence; and $\mathcal{V}$ specifies evaluation procedures, limits, responsibilities, and revision conditions. The record documents a proposed warrant; its existence does not establish that the comparison is correct.

The record in Definition 11.1 is intentionally broad enough to accommodate a mathematical proof, a laboratory comparison, a historical investigation, or a collaborative interpretive exercise. Its entries have different evidential standards. In a finite formal example, E may be a reproducible transition table and $\mathcal{V}$ a proof and exhaustive check. In an empirical kinetic study, E would include instruments, calibration, repeated observations and appropriate statistical analysis. In a participatory linguistic comparison, E may include annotated texts and documented interpretive judgments. These cases warrant different conclusions; the common record helps show where their claims become comparable and where they remain distinct.

Table 11 sets out a sequence of checks and the kinds of adverse result they should expose. Each stage should retain the outputs of specialist inquiry, including findings that make the G4 layer unnecessary for the declared task. The protocol can be implemented lightly for simple comparisons and in greater detail when consequences or uncertainties warrant it.

Table 11: Proposed Stages, Specialist Baselines, and Revision Triggers for Cross-System Comparison.

Evaluation stage	Evidence or specialist baseline	Revision trigger
Object and concept	Target definition, domain expertise, measurement-validity review	Different operational meanings for a required common construct
Mapping and invariance	Explicit relation, counterexample search, formal or numerical verification	Failure to preserve the selected relation on a justified domain
Feasibility and history	Inventory, constraints, provenance, competing trajectory models	An omitted dependency changes the admissible continuations
Identification and causation	Model set, uncertainty analysis, controls and possible interventions	Rival mechanisms yield opposite conclusions under existing evidence
Held-out comparison	Specialist workflow, benchmark tasks, uncertainty and resource costs	No credible added discrimination or excessive procedure burden
Institutional application	Authority record, participation, review, correction and legal context	Category use exceeds its warrant or blocks justified correction

The protocol should record failed and inconclusive comparisons alongside successful ones. A failed mapping can identify the role of a hidden resource or a contested concept; an inconclusive comparison can reveal a necessary measurement that remains unavailable. Preserving these outcomes prevents the framework from becoming an archive of selectively successful analogies. Reanalysis should be possible after changed instruments, new evidence, or an improved specialist model, with earlier versions of the comparative claim retained for inspection.

11.9 Revision Criteria and Boundaries of the Programme

The framework faces limitations of several kinds. *Formal limits* arise when proposed mappings fail to preserve selected relations or when composition requires incompatible regimes. *Epistemic limits* arise when observationally compatible models disagree about the comparison target. *Historical limits* arise when records of formation and dependency remain incomplete. *Practical limits* arise when an accurate comparison is too costly, intrusive, or weakly informative to justify its use. *Institutional limits* arise when a technically defensible category acquires authority over questions for which it lacks adequate evidential or ethical justification.

Three outcomes should lead to different revisions. Where a particular correspondence fails a well-designed discriminating test, the appropriate response is to restrict or withdraw that correspondence and identify the source of failure. Where multiple candidate models remain equally compatible with available evidence, the conclusion should remain partially identified or explicitly unresolved, with a research plan for acquiring relevant information. Where the framework's own comparative protocol yields no defensible gain over established specialist approaches, its claimed methodological contribution should be reduced to an optional way of documenting research assumptions. A finding that its categories systematically cause false confidence, conceal relevant differences, or impede correction would provide stronger grounds to revise or abandon the proposed procedure.

An apparent success also has limited implications. Preservation of one observable under an intervention is evidence for the specified correspondence under those conditions. It does not validate every structural property of the compared systems, all possible regimes, or a shared ontology. For the present paper, *cross-relational invariance* remains a task-relative claim whose object, preservation standard, and evidence must be declared. The findings of Sections 7 and 10 explain why uncertainty about models and the social authority of categories can persist even when a formal property is preserved exactly.

This section therefore establishes a programme of criticism rather than a completed external evaluation. Its constructive proposal is a disciplined sequence of comparative claims whose vulnerability to counterevidence is documented at the time each claim is made. Its success will depend on whether independent researchers find that the sequence improves relevant judgments under fair comparisons with domain-specific practice. Until such tests are undertaken, G4's scope remains a provisional philosophi-

cal and methodological synthesis. The final section draws together the conditional results and identifies a bounded agenda for subsequent research.

12 Conclusion

Comparability among heterogeneous generative systems concerns the justification of inferences across distinct descriptions, conditions, and histories. The paper has developed a task-relative account of comparative warrant, specifying objects, admissible conditions, partial mappings, selected preservation claims, evidence, and criteria for revision. This Conclusion brings together the conditional mathematical results, historical and interpretive distinctions, and institutional implications, while identifying the evaluations still needed. Its organizing commitment is to preserve informative partial comparisons together with their unresolved differences and the grounds on which their interpretation may change.

12.1 Comparative Warrant as a Situated Research Object

The paper's starting point is a distinction between describing a common feature and establishing an adequate relation between two systems. A biological circuit and a computational controller can both be represented by directed cycles. The selected diagram may preserve adjacency and reachability, while the systems differ in kinetic rates, substrates, responses to perturbation, and physical implementation. A shared label can likewise coordinate grammatical annotation across languages while leaving unresolved differences in linguistic interpretation and annotation practice. In each case, the warrant of the comparison is determined by the question being asked and by the information preserved for that question. Section 2 formalized this orientation through interpreted candidate-system descriptions and task-relative comparative adequacy.

Comparative warrant is therefore a situated research object rather than a property inferred directly from resemblance. It can be recorded, evaluated, contested, and revised. Its relevant dimensions include the selected target phenomena, the status of their representations, the conditions of observation, and the inferential transition from an observed correspondence to a substantive conclusion. These dimensions can change separately. Improved measurement might distinguish two configurations previously grouped together. A change in resources can alter joint feasibility while preserving a diagram of possible transformations. A new interpretive agreement may enable a local translation without eliminating other disagreements. An institutional change can affect which comparisons receive authority while leaving their formal content intact.

The approach provides a vocabulary for qualifying such changes. It distinguishes exact from approximate preservation, local from global scope, marginal from joint possibility, and descriptive similarity from identification of a generative mechanism. These distinctions have precedents in existing mathematical and philosophical research. Their coordinated deployment supplies a possible methodological benefit when an inquiry crosses disciplinary or institutional boundaries and requires several forms of justification at once. This is the paper's organizing contribution. It is a proposal about how comparative claims can be assessed, with practical usefulness awaiting evaluation against competent specialist alternatives.

12.2 Mathematical Preservation and Its Conditional Scope

The mathematical development makes explicit the structure of selected comparative claims. In Section 3, a partial mapping is declared on a specified domain, with preservation distinguished from reflection. Under appropriate assumptions, a bijective mapping that preserves and reflects the selected relations supports a local relational isomorphism. A many-to-one mapping supports an exact quotient only when the underlying relation is suitably saturated across the relevant fibres; the criterion is stated in Proposition 3.4. Forward preservation can also be composed under the declared domain and regime compatibility requirements of Proposition 3.5. These are established kinds of relational reasoning, organized here to prevent

an unexamined extension from partial correspondence to whole-system equivalence.

Section 4 further distinguishes the properties that transformations actually preserve. A commuting correspondence between two updates yields preservation of their finite iterations under the assumptions of Proposition 4.2. Selected conserved quantities can be transported when the relevant maps satisfy the conditions of Proposition 4.3. Bounded observational or weighting discrepancies support stable numerical comparisons only when a suitable margin remains, as illustrated by Proposition 4.4 and the institutional ranking result of Proposition 10.1. Such statements characterize precise preservation tasks. They leave open the empirical adequacy of the modeled update, observable, and measurement error.

The paper consequently treats an invariant as an explicitly indexed finding. The relevant index includes the representation, the chosen domain, the transformation class, and the claim the invariant is intended to support. Preservation of a graph property answers a different question from preservation of a stochastic response, a conserved mass balance, or an interpreted linguistic distinction. Distinct task indices can sometimes be coordinated, but their coordination requires justified connections between the corresponding descriptions. Exact mathematical results thereby become useful safeguards against interpretive extension beyond the selected property.

The formal results retain the provenance of their mathematical baselines. Isomorphisms, relation-compatible quotients, composition, conservative quantities, resource reachability, history-based congruence, and fibrewise identification are established forms of mathematical reasoning. G4 assembles them in an inquiry-specific warrant framework, while the explanatory or practical gain of such coordination remains to be demonstrated against disciplinary methods. Section 11.2 records the resulting proof and interpretation boundaries.

12.3 Generative Conditions, Provenance, and Incomplete Reconstruction

Comparisons of generative processes must account for the conditions under which their transformations are feasible. Section 5 examined joint prerequisites, consumable resources, shared capacities, and regime compatibility. A collection of outcomes that can each be reached from a given initial inventory may exceed what can be realized together. Conversely, reachability factorizes for independent systems under the restrictive assumptions of Proposition 5.3. These constructions offer contrasting reference cases. Shared resources, common institutional permissions, and ecological dependencies can defeat a comparison based on independent composition, whereas adequately separated subsystems may permit it. The existence and significance of such dependencies require domain-specific evidence.

Historical provenance adds another dimension. Systems with the same present description can retain different inventories, permissions, learned configurations, or latent response conditions. The analysis in Section 6 specifies a condition under which histories sharing an observable description have compatible observable continuations over finite sequences of actions. The result makes a limited predictive claim about a selected history domain. It permits historical traces to be omitted when the available evidence and the chosen predictive task justify doing so. Where that condition fails, recovering further aspects of formation can improve comparison. Explanatory claims about why a particular system formed as it did require a distinct historical or causal investigation.

Section 7 examined the case in which the underlying structure is itself incompletely observed. Candidate mechanisms, admissible correspondence maps, and historical reconstructions can remain compatible with the same record. Proposition 7.2 identifies the elementary condition under which a comparative target is constant on observationally indistinguishable candidates, while Proposition 7.3 describes the narrowing of partially identified target sets when compatible candidates are eliminated by additional evidence. A well-defined comparison can therefore remain unresolved with the available observations. This is an epistemic result rather than a defect that can be repaired through more confident terminology.

Taken together, material feasibility, provenance, and identifiability define a discipline of restraint. They indicate the information needed before extending a comparative statement from one observation, inventory, or history to another. They also suggest discriminating research designs: alter a resource constraint, recover a relevant historical trace, compare responses to controlled interventions, or intro-

duce measurements that separate otherwise compatible candidate models. Where such procedures are unavailable, the resulting limitation should remain part of the reported conclusion.

12.4 Reciprocal Interpretation and Responsibility for Categories

Relational correspondence can involve meanings that are revised through interaction. Section 8 considered partial reciprocal translations, the maintenance of selected distinctions under a round trip, and the preservation of comparisons calculated from shared probe signatures. The associated propositions concern bounded domains and declared tests. Their validity depends on the selected equivalence and calibration rules. They establish the conditions for retaining certain distinctions across translation, while the wider question of mutual understanding remains sensitive to history, use, practices of correction, and disagreements about relevant purposes.

An additional issue arises when comparative categories influence decisions about other participants. The possibility of measuring, classifying, or ranking something is distinct from the authority to make that classification consequential. Section 10 examined how metric design, access to evidence, standing, and revision procedures shape this authority. Its bounded score-stability result identifies circumstances in which a ranking survives a declared perturbation; stability of that ranking remains a mathematical property of the adopted metric. Questions of participation, accountability, legality, and fairness require independent judgment and evidence. The same distinction applies to comparisons involving unfamiliar or hypothetical forms of agency, for which evidence about capacities, experience, or moral standing may be especially incomplete.

The practical implication is a requirement for visible custody of comparative categories. Investigators should identify who selected a measure, which affected perspectives entered its design, what information remains inaccessible, and how material errors can be challenged. A useful audit should also record the effects of correction: whether a disputed classification was changed, whether associated inferences were revised, and whether consequential applications reflected the corrected record. Institutional processes with these properties may support more responsible comparison, although their effectiveness varies by setting and must be investigated empirically. Epistemic accountability thereby enters the research programme as a separate, observable issue rather than as a legitimacy claim automatically derived from structural preservation.

12.5 Comparative Cases and Specialist Explanatory Standards

The cases examined in Section 9 illustrate how a comparative warrant may be tested against the explanatory standards of the relevant field. A common directed cycle between a synthetic regulatory circuit and a computational controller can support a narrow structural claim, while kinetic predictions require additional physical information. A resemblance between an ecological food-web graph and a resource-flow model can conceal differences in biomass, renewal, and constraints on simultaneous consumption. A common multilingual annotation scheme can assist computational comparison while remaining open to language-specific interpretation and revision. Different historical formation paths can produce equivalent present records alongside unequal capacities for further action.

Established discipline-specific methods remain the appropriate baselines. Graph methods evaluate reachability and transition-system methods address behavioral correspondence. Resource-sensitive models represent joint prerequisites, while stochastic approaches evaluate kinetics and uncertainty. Measurement theory examines the relation between operational indicators and intended constructs. Historical and interpretive inquiry investigates provenance and meaning. G4's cross-relational proposal gains credibility when it accurately identifies the limits of each translation and supports better coordination between these methods. Its credibility would be diminished by a claim of universality that overlooks the kinds of knowledge these methods already provide.

12.6 Evaluation Priorities and Conditions for Revision

The programme's next stage is comparative evaluation on predetermined tasks. Section 11 introduced a Task-Relative Comparative Warrant Record in Definition 11.1, together with specialist baselines, competing hypotheses, and explicit revision triggers. A defensible study would prepare multiple kinds of comparative problems, including exact finite formal models, cases with competing latent mechanisms, tasks involving consumptive resources, cross-language interpretive problems, and institutional settings in which a category informs a consequential decision. The task battery should include deliberately difficult negative cases, and its scoring criteria should be fixed before the methods are compared.

A first evaluation could test *error detection*: whether a structured warrant record helps investigators identify omitted resources, invalid mapping domains, incompatible histories, or unjustified causal claims. A second could test *decision calibration*: whether investigators express appropriate uncertainty when the candidate-model set supports only partial identification. A third could test *revision quality*: whether new evidence leads to a traceable correction of the asserted correspondence and its practical use. In a collaborative setting, the record could be assessed for the accessibility of reasons, the opportunity to challenge a category, and the burden of maintaining the procedure. These are distinct research outcomes with different instruments and possible harms.

The evaluation should also expose two methodological choices capable of changing an apparent result without new evidence. Section 7.6 shows how enlarging the family of admissible models can reopen a previously point-identified comparison. Section 11.6 shows why task-specific evaluation losses require justified scales and weights before they can be aggregated. These decisions should remain visible in any report of the framework's performance. Their sensitivity forms part of the comparison's warranted scope rather than a secondary editorial detail.

The programme also requires clear conditions for contraction. If a purported invariant fails under an admissible transformation, its domain should be restricted or its preservation rule revised. If rival models compatible with the evidence support opposite claims, the appropriate reported result is partial identification or an unresolved target. If a domain-specific method already provides equally accurate, interpretable, and accountable judgments with lower overhead, the G4 layer may reasonably remain optional. If the chosen category systematically suppresses relevant differences or excludes justified challenges, the institutional procedure itself requires substantial revision. Recording these outcomes gives the research programme a way to learn from its failures.

12.7 Open Ontological Questions and the Scope of Synthesis

The inquiry leaves open the ontological significance of similarities that survive many distinct representations. A property may remain stable across a variety of measurement choices and transformations, and that stability can support confidence in its scientific usefulness. Whether it indicates a physically real common structure, a robust effective description, or the success of a restricted modeling practice depends on further arguments about the relation between the representation and its target. Structural realism, pragmatist approaches to representation, and relational ontology provide different resources for that argument. G4 uses these traditions as interlocutors while suspending an exclusive ontological conclusion.

The limits of mutual translation also deserve sustained philosophical inquiry. A partial correspondence can preserve a useful operational relation while leaving parts of a conceptual system untranslatable or disputed. Such residuals may become resources for new research questions or jointly revised concepts. Their significance is especially evident when interacting participants possess different forms of knowledge, different stakes in the comparison, or different capacities to contest an inherited vocabulary. The prospect of revising comparative categories together is philosophically important, but it cannot substitute for evidence about actual practices of participation and change.

G4's scope therefore remains bounded. G1 examines the ontological status of derived generative order; G2 examines political-economic differences among heterogeneous generative-relational subjects; G3 investigates evolutionary formation of norms among such subjects. The present inquiry addresses the warrants for comparing systems whose configurations, transformations, descriptions, and histories

may differ. It can inform those neighboring projects by identifying when an analogy carries a defensible invariant and when the comparison requires additional evidence. The distinct research objects and standards of assessment of the four projects should be preserved.

12.8 Concluding Position on Heterogeneous Comparability

The outcome is a conditional account of heterogeneous comparability. Selected relations, observables, and transformations support comparison when their domains, preservation criteria, and evidential standing are stated. Material dependencies, historical trajectories, interpretation, and institutional authority delimit the conclusions that follow. Mathematical constructions identify precise preservation claims; the proposed warrant record supports their disclosure, assessment, and revision.

G4’s methodological proposal concerns the intelligibility and accountability of cross-system inference. Its formal results rest on established techniques; its illustrative cases and evaluation protocol remain preliminary. Independent comparative studies will determine whether this organization improves research practice beyond competent specialist alternatives. The present contribution is a transparent and revisable way to distinguish warranted local correspondence from claims whose evidence or interpretation remains unsettled.

Research Ethics and Source Provenance

This independent unreviewed preprint presents all twelve substantive sections of the paper. It makes no claim to independent empirical verification or priority over established mathematics, philosophy of science, or comparative methodology. Bibliographic records and their recorded access limits are documented in the source-verification ledger; many references have been examined at the metadata, abstract, or reference-article level, with comprehensive full-text source review outstanding. Citations are used only for the limited claims stated in the drafted sections. AI assisted with research dialogue, English drafting, source discovery, and LaTeX preparation. The human author remains responsible for all scholarly judgments, accuracy, and public release decisions.

References

- [1] Robert Adcock and David Collier. “Measurement Validity: A Shared Standard for Qualitative and Quantitative Research”. In: *American Political Science Review* 95.3 (2001), pp. 529–546. DOI: 10.1017/S0003055401003100.
- [2] Elizabeth S. Allman, Catherine Matias, and John A. Rhodes. “Identifiability of Parameters in Latent Structure Models with Many Observed Variables”. In: *The Annals of Statistics* 37.6A (2009), pp. 3099–3132. DOI: 10.1214/09-AOS689.
- [3] W. Brian Arthur. “Competing Technologies, Increasing Returns, and Lock-In by Historical Events”. In: *The Economic Journal* 99.394 (1989), pp. 116–131. DOI: 10.2307/2234208.
- [4] John C. Baez and Jade Master. “Open Petri Nets”. In: *Mathematical Structures in Computer Science* 30.3 (2020). Compositional operational and reachability semantics for open Petri nets, pp. 314–341. DOI: 10.1017/S0960129520000043.
- [5] Geoffrey C. Bowker and Susan Leigh Star. *Sorting Things Out: Classification and Its Consequences*. MIT Press, 1999. DOI: 10.7551/mitpress/6352.001.0001.
- [6] Katherine Brading, Elena Castellani, and Nicholas Teh. “Symmetry and Symmetry Breaking”. In: *The Stanford Encyclopedia of Philosophy*. Ed. by Edward N. Zalta and Uri Nodelman. Fall 2023. Substantive revision 1 August 2023. Metaphysics Research Lab, Stanford University, 2023. URL: <https://plato.stanford.edu/archives/fall2023/entries/symmetry-breaking/>.
- [7] Donald Davidson. “On the Very Idea of a Conceptual Scheme”. In: *Proceedings and Addresses of the American Philosophical Association* 47 (1974), pp. 5–20. DOI: 10.2307/3129898.
- [8] Jennifer A. Dunne, Richard J. Williams, and Neo D. Martinez. “Food-Web Structure and Network Theory: The Role of Connectance and Size”. In: *Proceedings of the National Academy of Sciences* 99.20 (2002), pp. 12917–12922. DOI: 10.1073/pnas.192407699.

- [9] Michael B. Elowitz and Stanislas Leibler. “A Synthetic Oscillatory Network of Transcriptional Regulators”. In: *Nature* 403 (2000), pp. 335–338. DOI: 10.1038/35002125.
- [10] Wendy Nelson Espeland and Mitchell L. Stevens. “Commensuration as a Social Process”. In: *Annual Review of Sociology* 24 (1998), pp. 313–343. DOI: 10.1146/annurev.soc.24.1.313.
- [11] Brendan Fong and David I. Spivak. *An Invitation to Applied Category Theory: Seven Sketches in Compositionality*. Cambridge University Press, 2019. DOI: 10.1017/9781108668804.
- [12] Miranda Fricker. *Epistemic Injustice: Power and the Ethics of Knowing*. Publisher book and chapter abstracts consulted for testimonial and hermeneutical injustice. Oxford University Press, 2007. DOI: 10.1093/acprof:oso/9780198237907.001.0001.
- [13] Roman Frigg and James Nguyen. “Scientific Representation”. In: *The Stanford Encyclopedia of Philosophy*. Substantive revision 20 May 2026. 2026. URL: <https://plato.stanford.edu/entries/scientific-representation/>.
- [14] Jean-Yves Girard. “Linear Logic”. In: *Theoretical Computer Science* 50.1 (1987), pp. 1–101. DOI: 10.1016/0304-3975(87)90045-4.
- [15] Miguel A. Hernán and James M. Robins. *Causal Inference: What If*. Boca Raton: Chapman & Hall/CRC, 2020. URL: <https://miguelhernan.org/whatifbook>.
- [16] Wilfrid Hodges. *Model Theory*. Cambridge University Press, 1993. DOI: 10.1017/CB09780511551574.
- [17] C. S. Holling. “Resilience and Stability of Ecological Systems”. In: *Annual Review of Ecology and Systematics* 4 (1973), pp. 1–23. DOI: 10.1146/annurev.es.04.110173.000245.
- [18] Jari P. Kaipio and Erkki Somersalo. *Statistical and Computational Inverse Problems*. Applied Mathematical Sciences 160. Springer, 2005. DOI: 10.1007/b138659.
- [19] Helen E. Longino. *Science as Social Knowledge: Values and Objectivity in Scientific Inquiry*. Conceptual background on evidence, values and social inquiry; interpretation limited to publisher summary. Princeton University Press, 1990. ISBN: 9780691020518. DOI: 10.2307/j.ctvx5wbfz.
- [20] Saunders Mac Lane. *Categories for the Working Mathematician*. 2nd ed. Graduate Texts in Mathematics 5. Springer, 1998. DOI: 10.1007/978-1-4757-4721-8.
- [21] Charles F. Manski. *Partial Identification of Probability Distributions*. Springer Series in Statistics. Springer, 2003. DOI: 10.1007/b97478.
- [22] Sally Engle Merry. *The Seductions of Quantification: Measuring Human Rights, Gender Violence, and Sex Trafficking*. University of Chicago Press, 2016. ISBN: 9780226261287.
- [23] Joakim Nivre et al. “Universal Dependencies v2: An Evergrowing Multilingual Treebank Collection”. In: *Proceedings of the Twelfth Language Resources and Evaluation Conference*. European Language Resources Association, 2020, pp. 4034–4043. URL: <https://aclanthology.org/2020.lrec-1.497/>.
- [24] Eric Oberheim. “The Incommensurability of Scientific Theories”. In: *The Stanford Encyclopedia of Philosophy*. Substantive revision 7 May 2024. 2024. URL: <https://plato.stanford.edu/entries/incommensurability/>.
- [25] Paul Pierson. “Increasing Returns, Path Dependence, and the Study of Politics”. In: *American Political Science Review* 94.2 (2000), pp. 251–267. DOI: 10.2307/2586011.
- [26] Theodore M. Porter. *Trust in Numbers: The Pursuit of Objectivity in Science and Public Life*. Princeton University Press, 1995. ISBN: 9780691037769.
- [27] Steven P. Reise, Keith F. Widaman, and Robin H. Pugh. “Confirmatory factor analysis and item response theory: two approaches for exploring measurement invariance”. In: *Psychological Bulletin* 114.3 (1993), pp. 552–566. DOI: 10.1037/0033-2909.114.3.552.
- [28] Davide Sangiorgi. *Introduction to Bisimulation and Coinduction*. Print publication 2011; electronic edition 2012. Cambridge University Press, 2011. DOI: 10.1017/CB09780511777110.
- [29] Jason Seawright and John Gerring. “Case Selection Techniques in Case Study Research: A Menu of Qualitative and Quantitative Options”. In: *Political Research Quarterly* 61.2 (2008), pp. 294–308. DOI: 10.1177/1065912907313077.
- [30] Cosma Rohilla Shalizi and James P. Crutchfield. “Computational Mechanics: Pattern and Prediction, Structure and Simplicity”. In: *Journal of Statistical Physics* 104.3-4 (2001), pp. 817–879. DOI: 10.1023/A:1010388907793.
- [31] Robert J. Vandenberg and Charles E. Lance. “A Review and Synthesis of the Measurement Invariance Literature: Suggestions, Practices, and Recommendations for Organizational Research”. In: *Organizational Research Methods* 3.1 (2000), pp. 4–70. DOI: 10.1177/109442810031002.