Human and Artificial Reality: Generative Relations and the Formation of Interpretation

Wanhong HUANG
huangwanhong@serendip.ngo

Abstract

Comparisons between human and artificial reality often move between existence, observable performance, social consequence, and subjective experience without specifying the inference connecting them. This discussion paper distinguishes ontological, relational, and manifest dimensions of reality attribution and examines the historical formation of the systems that interpret them. It argues that comparison requires an explicit domain, observation practice, temporal horizon, and class of interventions. The progressive strengthening of indistinguishability assumptions changes which differences remain evidentially available; further arguments determine their explanatory or normative significance. Embodiment and historical dependence enter this account as potentially consequential conditions whose importance requires specification. A completed exploratory pilot collected 24 actual agent responses, all selecting equality under their stated criteria while leaving consciousness and ultimate nature unresolved. The result is limited by a possible existential ceiling and single-platform design. Separate field simulations and exact calculations distinguish projected agreement, dynamical conjugacy, orbit equivalence, shared attractors, action values, and boundary state sums. These examples clarify failures of unrestricted inference without validating a theory of artificial subjectivity. The proposed contribution is a disciplined connection between relational explanation, situated interpretation, and the scope of equivalence claims, open to correction and comparison with established traditions.

Keywords: reality attribution; artificial intelligence; relational interpretation; embodiment; dynamical equivalence; field models.

Research Paper Note

This philosophical discussion paper develops conceptual distinctions and examines them through hypothetical comparisons, a small live-agent pilot, and synthetic mathematical constructions. The studies do not establish artificial consciousness or the metaphysical identity of human and artificial systems.

The project welcomes objections, corrections, competing formulations, and identification of related disciplinary work. It claims no priority in relationality, situated interpretation, the emergence of meaning, or classification feedback. Initial neighbouring sources include perspectival realism, participatory sense-making, social-relational robot ethics, and the genealogy of AI personification [8, 9, 14, 18]. The accompanying source ledger states the access and claim limits of this initial review; the literature review develops a selective comparison with these antecedents.

Responsible Use and Rights Reservation

This section records requested scholarly conduct and the scope of retained rights. The scholarly requests below accompany the preprint licence and introduce no additional condition on permissions granted by that licence.

Good-faith criticism, correction, independent inquiry, and responsible reuse are welcome. The conceptual distinctions should be evaluated with their assumptions and limits intact. In particular, observed

performance and social effects require separate arguments before supporting claims about consciousness, moral status, or a hierarchy of reality.

Rights outside the preprint licence remain with their respective holders. A request for responsible scholarly use introduces no additional restriction to that licence. No exclusive entitlement to the discussed ideas, concepts, or independent formulations is asserted.

Notices

Preprint status. This is a discussion preprint. Independent human peer review has not been completed. The arguments, interpretations, and proposed applications remain open to correction.

Preprint licence. The original material in this preprint is made available under Creative Commons Attribution–NonCommercial 4.0 International (CC BY-NC 4.0), https://creativecommons.org/licenses/by-nc/4.0/, to the extent that copyright or related rights apply and are held by the contributors. Reuse should follow the licence's attribution and noncommercial conditions, retain the licence notice, and identify changes. Separately identified third-party material remains subject to its own rights and permissions. Material in the public domain and uses permitted by applicable exceptions remain unaffected. The official licence controls its terms.

AI-use disclosure. The supplied exploratory human–AI conversation informed the research questions. An OpenAI coding assistant subsequently prepared titles, outlines, initial source checks, methodological criticisms, experimental materials, and the complete manuscript drafts, expanded revisions, literature review, and integrated appendices. AI assistance extended to substantive drafting and analysis, beyond language editing or formatting. Twenty-four freshly forked agents returned actual responses in a single-platform exploratory pilot. Raw responses, prompts, dispatches, event records, and analysis are preserved. The exact backend checkpoint and sampling settings were unavailable. Separate Python experiments examined a finite transition system and a phenomenological spatial field model, with the latter's parameters independent of the agent responses. A further assistant reviewed the methods within the same platform; this review is distinct from independent human peer review. Subsequent computational studies examined field conjugacy and observation, mobility-dependent variational dynamics, and a finite-group lattice gauge state sum with a dual spin-foam-type representation. These studies retain their own comparison maps, boundaries, and normalization; they supply no gravitational or subjective interpretation of the experimental agents. The assistant's output has not been represented as an independently approved human manuscript. No human-participant study, independent human coding, or empirical validation of a developmental mechanism has been completed.

Human attribution. The author is Wanhong HUANG, contact `huangwanhong@serendip.ngo`. AI systems are disclosed as research and drafting tools. The author attribution identifies the human contributor and does not represent independent peer review. Earlier sources and PDFs are preserved in the project version archive.

1 Introduction

This paper examines the grounds on which human and artificial systems are compared as real, meaningful, and historically consequential. Its purpose is to distinguish the claims involved in such comparisons and to develop a relational account of the formation of interpretation. The inquiry combines conceptual analysis, controlled hypothetical comparisons, an exploratory study of actual agent responses, and mathematical constructions. Each method addresses a different part of the argument, and the evidential boundaries between them remain explicit.

A human interlocutor and an artificial interlocutor can produce similar utterances while differing in their histories, material organization, social positions, and capacities. A judgment that one is more real can refer to existence, authenticity, subjective experience, causal efficacy, or recognition within a practice. The ambiguity becomes consequential when an observation about one of these dimensions is treated as sufficient evidence for another. Comparable dialogue, for example, establishes a particular performance comparison. Its implications for consciousness, personal identity, or moral standing depend on additional premises. Conversely, a difference of origin requires an argument specifying the respect in which that difference matters.

The problem extends to the interpreter. Criteria of intelligibility and credibility are acquired through bodies, histories, institutions, and repeated interactions. An interpreter's perspective is therefore part of the situation to be explained. This observation creates a difficulty for comparisons that silently treat one historically formed standpoint as a neutral measure of every possible form of intelligence. It also creates a difficulty for unrestricted relativism: acknowledging the formation of criteria does not settle which claims have adequate evidence or which practices are justifiable. The paper develops conditions under which perspectives can be situated and still answerable to reasons, consequences, and correction.

The central proposal has three components. First, ontological commitment, relational efficacy, and manifest appearance should be distinguished within a comparison. Second, the histories that form an interpretive system should enter the explanation when they change its organization, dispositions, or response to a relevant intervention. Third, an equivalence claim should identify its objects, observation conditions, admissible transformations, and temporal scope. The phrase *generative relations* names the proposed explanatory emphasis on the processes through which capacities, expectations, and patterns of recognition develop. It carries no claim to conceptual priority or to a completed theory of mind.

The methodological contribution is an articulated set of inferential tests. The live-agent pilot examines how prompted respondents formulate a criterion under stipulated comparisons. It provides a small corpus of actual responses, with a design that also exposes a possible ceiling in the question asked. The spatial field construction demonstrates how equal accessible histories can coexist with retained differences that a later probe reveals. The comparative field, action, and boundary calculations then identify distinct senses of sameness. These constructions provide counterexamples to unrestricted inferences; their variables have no established measurement correspondence to human or artificial subjectivity.

The paper proceeds from the literature review in Section 2 to the scope of reality attribution in Section 3. Sections 4–8 develop the conceptual distinctions, historical formation, observer displacement, and controlled comparisons. Section 9 relates these arguments to the completed studies. Section 10 addresses philosophical objections, and Section 11 states the resulting research commitments. The appendices incorporate the methods, complete response corpus, field derivations, and comparative calculations within this paper. The companion paper, *Generative Relations in Legal Judgment: Evidence, Historical Trajectories, and Norm-Relevant Similarity*, develops a separate normative inquiry; its legal proposals are not treated as consequences established by the present experiments.

2 Literature Review

This section locates the inquiry among established approaches to machine comparison, relational cognition, situated knowledge, classification, and formal equivalence. The review is thematic and selective. It identifies antecedents relevant to the argument and specifies the narrower role each plays here. Bibliographic details and the scope of the cited claims were checked online before insertion; some sources were accessible only through publisher abstracts or selected original passages. The review therefore supports a provisional positioning of the paper, with further disciplinary comparison remaining part of its research agenda.

2.1 Machine Comparison and Operational Criteria

Turing's discussion of machine intelligence provides a foundational precedent for replacing an unrestricted question with a specified comparison procedure. The opening of *Computing Machinery and Intelligence* introduces the imitation game while examining the difficulty of defining the initial question through ordinary usage [20]. The present inquiry inherits the methodological concern with what a procedure licenses an investigator to infer. Its target is reality attribution, whose possible meanings extend beyond the assessment of intelligent performance. The live-agent vignette is consequently a study of articulated judgments about stipulated cases, with no claim to constitute an implementation of Turing's procedure or a test of consciousness.

A specified test can improve a comparison while preserving uncertainty about its interpretation. The choice of interlocutors, permitted observations, time horizon, and interventions determines which differences are available to the investigator. A successful match may then bear on a capacity, a response distribution, or a practical role. The movement from that result to a claim about the system's constitution remains a further argumentative step. This paper uses the distinction to keep operational achievements and metaphysical interpretations separately assessable.

2.2 Coupled Cognition and Participatory Sense-Making

Clark and Chalmers argue that active coupling with environmental resources can form part of a cognitive process. Their discussion treats the organization of an ongoing coupled system as consequential and explicitly distinguishes cognition from consciousness [7]. This antecedent supports examination of the boundary selected around an interpreter. It does not establish that every causal support belongs to a mind, that consciousness extends wherever cognition does, or that an artificial participant acquires moral standing through coupling alone. Those stronger claims require their own criteria.

De Jaegher and Di Paolo's account of participatory sense-making gives social interaction a constitutive role in the discussion of meaning [9]. The relevance here is the possibility that a pattern of interaction helps explain what participants can make of a situation. The present use remains limited to that relational orientation; the three-field equations below are independently stipulated and are not offered as a mathematical derivation of enactive theory.

These approaches motivate a distinction between a momentary internal state, the organization of a coupled process, and the developmental conditions that make such organization possible. A change in an input can alter a response without reorganizing the interpreter's capacities. A change in available practices, categories, or resources may require a more extensive account. The conceptual discussion develops this distinction explicitly because the retained-history simulation demonstrates state dependence under fixed equations. A stronger developmental thesis must identify which structures change, how they change, and how the change can be observed.

2.3 Perspectives, Reflexivity, and Classification

Massimi's *Perspectival Realism* provides a neighbouring framework for discussing knowledge through perspectives while retaining realist commitments [14]. The connection concerns the compatibility of situated inquiry with answerability to a world. This paper does not attempt a reconstruction of that book's complete epistemology. Its narrower question is how a comparison can acknowledge an interpreter's position while preserving publicly inspectable evidence and conditions of correction.

Bourdieu and Wacquant's *An Invitation to Reflexive Sociology* identifies methodological relationalism, the logic of fields, and reflexivity as established topics in social theory [6]. This ancestry matters both terminologically and methodologically. The present mathematical use of a field is defined by a domain and values assigned across it. No identification with a sociological field follows from the shared word, and no derivation of the simulation from Bourdieu's theory is claimed. A substantive connection would require an account of positions, relations, practices, and their measurement beyond the publisher-level material consulted here.

Hacking's discussion of making up people supplies an antecedent for examining feedback between classifications and those classified [12]. Its relevance is the possibility that a description participates in the development of the situation it describes. Such feedback differs from an assertion that classification creates every property of its object. In the present inquiry, an attribution of artificial competence or personhood may affect interaction, access, and expectations even while important questions about the attributed capacity remain open.

The resulting methodological obligation is reflexive without being unrestricted. An investigator should make the selection of a comparison domain available for criticism, examine exclusions that matter to the stated claim, and preserve the distinction between a social consequence and the truth of the description producing it. A successful intervention can alter the relations under study; this alteration can itself become evidence, provided the relevant before-and-after conditions are specified.

2.4 Relational Recognition and Artificial Personification

Coeckelbergh's social-relational approach to robot rights is an antecedent for considering relations in moral consideration [8]. Romele's work on relational personhood and genealogies of illusion supplies a complementary direction for investigating the conditions of AI personification [18]. These references identify an existing debate rather than a consensus adopted here. Recognition can have practical effects, while its warrant and consequences remain open to criticism.

The distinction matters to both sympathetic and critical interpretations of artificial systems. A user may form a consequential attachment under conditions of uncertainty. That attachment can warrant study of dependence, responsibility, and design even when the system's subjectivity remains unresolved. Equally, a persuasive appearance of reciprocity can be produced under asymmetric control over information and interaction. The paper's account of relational reality must therefore include power and the conditions under which recognition develops. Describing a relation as consequential does not determine whether it should be encouraged, regulated, or withdrawn.

2.5 Field Dynamics and Comparison Structures

Neural-field and population-response models provide established mathematical precedents for studying spatial coupling and nonlinear response [2, 22]. Their citation acknowledges mathematical ancestry; their neuronal interpretations do not supply empirical support for the social meanings assigned to the present variables. The field appendix states its own assumptions and derives its own contraction and observation results.

The comparison of dynamical systems also has a substantial prior literature. Willems's behavioural perspective directs attention to admissible behaviour, while Girard and Pappas develop metrics for approximation between dynamical systems [11, 21]. These precedents help locate the distinction between a system representation, its accessible outputs, and an approximate match under specified conditions. The examples here assemble familiar types of comparison around an interpretive problem. They introduce neither a general new equivalence theorem nor a metric with unrestricted philosophical significance.

The variational and boundary examples broaden the structures under examination. Gradient-flow and large-deviation sources distinguish ingredients such as energy, mobility, and path functionals [1, 16]. Spin-foam and coarse-graining sources establish the importance of boundary data, amplitudes, and consistency under changes of description [3, 17]. The finite-group calculation included here has an explicit dual state-sum representation. Its limited algebraic relationship to those methods supplies no claim about gravitational physics, consciousness, or the physical realization of an interpreter.

2.6 Position of the Present Inquiry

The proposed contribution is a connection among these questions under declared inferential constraints. Existing traditions already address relationality, situated interpretation, classification feedback, operational comparison, and mathematical equivalence. The paper organizes these concerns around the attri-

bution of human and artificial reality, then tests the reach of selected inferences with worked constructions and an inspectable response corpus. Its adequacy depends on the clarity and usefulness of that organization, the correctness of the examples, and the quality of its responses to objections.

The empirical component occupies a narrower position. Research on persona prompting provides reasons to examine the effects and limits of instructed standpoints [13, 23]. A role instruction supplies a condition of response generation; it does not create a human participant or establish possession of the standpoint described. The pilot therefore documents what the responding agents said under the saved instructions. A general theory of human interpretation, artificial experience, or cross-platform judgment requires evidence beyond that corpus.

3 Human–AI Comparison and the Scope of Reality Attribution

This paper examines the grounds on which a human or artificial system is described as real and the conditions under which a difference between them becomes significant. Its method combines conceptual distinctions, controlled thought experiments, and limited computational examples. The discussion proceeds from the objects of reality attribution to the formation of interpreters, then considers increasingly demanding comparisons between systems. The central proposal is that a defensible comparison identifies both its subject matter and the relations through which evidence becomes available.

An artificial conversational system, a fictional character represented by that system, and a user's belief in the character have different conditions of existence. A working computer can produce a description of an absent person. The description can alter a user's conduct even when the purported person has never existed. An inquiry that asks whether the interaction is real therefore needs to distinguish the running process, the represented individual, the user's response, and any institutional consequences. The practical seriousness of an interaction supplies evidence about some of these objects while leaving others unresolved.

The expression "more real" intensifies this problem. It can indicate independent existence, physical constitution, authenticity of origin, persistence, experiential immediacy, or recognized social standing. These criteria sometimes support comparisons, yet their ordering requires an argument. A biological organism has a biological developmental history. An engineered system has a history of manufacture, training, operation, and maintenance. Describing the difference establishes a distinction in origin. A further premise is required to turn origin into a general ranking of reality, or to make that ranking decisive for every subsequent question.

The argument developed here retains ordinary commitments to bodies, artifacts, events, and evidence. It selectively suspends questions concerning ultimate substance or the sufficient conditions of consciousness where the proposed comparison supplies inadequate grounds for deciding them. This selective suspension permits investigation of relationships and appearances while preserving the possibility that further ontological commitments will matter. A conclusion about comparable performance consequently carries a narrower scope than a conclusion about comparable experience.

The paper also considers the historical formation of comparison itself. An interpreter acquires categories, expectations, habits of attention, and standards of relevance through processes that can become objects of inquiry. The same event may become a signal of danger, a routine instruction, or an irrelevant disturbance within differently organized histories. Explaining these differences requires attention to the conditions under which interpretation develops and to the evidence that could challenge it. A historical explanation of an interpretation and a justification of its truth remain distinguishable tasks.

Several established traditions already investigate aspects of this terrain. Massimi's perspectival realism connects scientific knowledge with situated perspectives while sustaining realist commitments [14]. De Jaegher and Di Paolo develop participatory sense-making as an account of social cognition [9]. Coeckelbergh proposes social relations as grounds for moral consideration of robots [8]. Romele's recent work addresses genealogies of AI personification through media [18]. These references identify substantive neighbouring inquiries. The present proposal concerns the scope of comparison across reality attribution, historical interpretation, and formal equivalence; a comprehensive assessment of its contribution

requires closer engagement with these traditions.

The resulting position is deliberately conditional. When two systems agree under a specified observation practice, that agreement warrants a conclusion within the practice's scope. When a claimed difference concerns an excluded dimension, its relevance requires a bridge from that dimension to the question under examination. When a decision involves interests, rights, or protection, explicit normative premises must accompany the descriptive comparison. The following sections develop these requirements and examine their limits.

4 Ontological, Relational, and Manifest Dimensions

This section distinguishes dimensions of reality attribution through examples of existence, causal involvement, and observed appearance. The distinctions organize questions about a common world; they do not divide objects into exclusive classes or establish degrees of fundamental worth. Their purpose is to make transitions between claims available for criticism.

4.1 Ontological Commitment and Selective Suspension

Ontological inquiry concerns what exists and what kinds of entities or processes a description presupposes. The modest commitment that an event occurred differs from a comprehensive account of its ultimate constitution. A record of a conversation presupposes a recording process and some relation between the record and an event. Its evidential use also depends on questions of authenticity and alteration. These ordinary commitments already constrain inquiry, even when a researcher suspends judgment about the metaphysics of mind.

Selective suspension should identify its target. In a comparison of response timing, it may be reasonable to leave consciousness unsettled. In a comparison of vulnerability to pain, the conditions of experience may become central. A claim can therefore be bracketed for one purpose and indispensable for another. The resulting division of labour is provisional and accountable to the problem being investigated. It supplies no general exemption from ontological argument.

Dependence also requires specification. A software process depends on hardware, power, and organizational support. A human life depends on bodily processes and environmental conditions. Dependence establishes connections that can explain persistence or failure; it supplies an incomplete basis for a general hierarchy. Some dependent entities have conditions of existence involving conventions, as institutional offices do. Others depend causally on their surroundings without being constituted by recognition. Treating every dependence as the same relation would obscure these differences.

4.2 Relational Reality and Historical Efficacy

Relational inquiry concerns an entity's participation in processes through which other conditions are sustained or transformed. A relationship can distribute resources, stabilize expectations, enable action, or make an outcome more likely. Its description should identify participants, direction, duration, and the mechanism attributed to their connection. The phrase "generative relation" will refer to a relation considered in this productive and historical capacity.

Historical efficacy includes brief and unsuccessful events. A short interruption can prevent a transaction; an abandoned instruction can change later expectations. Persistence and robustness are useful evidence for an enduring pattern, while transience remains compatible with consequential existence. Accordingly, the paper does not use stability as a definition of reality. Stability concerns a specified pattern under specified disturbances, and its importance varies with the question.

Relations also have boundaries of attribution. If a fabricated message causes a recipient to transfer funds, the message, the belief, and the transfer participate in a real causal sequence. That sequence supplies no independent confirmation of the fabricated sender. The effectiveness of a representation and

the existence of its referent require distinct evidence. Keeping those evidential questions separate allows the account to recognize deception without denying its consequences.

4.3 Manifest Reality and Observation Conditions

Manifest inquiry concerns what is accessible through a particular encounter or observation procedure. A transcript, a visual display, an institutional record, and an embodied interaction expose different aspects of a process. Manifestation therefore includes instruments, selection rules, time resolution, and interpretive categories. These conditions are parts of the account of access, even when their effects are small enough to neglect for a particular purpose.

An observation can be accurate and incomplete. Two conversations might share every recorded word while differing in unrecorded threats, expectations, or available alternatives. Conversely, different words can perform a similar role when their recipients have learned comparable conventions. A claim of manifest similarity should identify the features compared and the features left outside the record. Exclusion may be justified by an inquiry's purpose, but excluded information cannot subsequently be treated as evidence of agreement.

Manifest reality is consequently broader than illusion and narrower than a complete inventory of the system. The recorded acceptance of an instruction may be entirely authentic while leaving the conditions of acceptance unresolved. The interpretation of that record can improve through additional evidence. This possibility of correction connects manifestation with relational and ontological inquiry: observations support claims about processes, and claims about processes motivate further observations.

4.4 Representations, Referents, and Institutional Effects

The relation between representation and institution clarifies how these dimensions overlap. Consider a fictional organization that recognizes a digital certificate as evidence of membership. A valid certificate can have an institutional effect because the organization has procedures connecting its issuance to membership. A forged certificate can induce similar conduct temporarily, although it fails the procedure's validity conditions. The visual similarity of the certificates, the causal consequences of their use, and their institutional status are independently assessable.

The example also identifies a limit to recognition-based accounts. Widespread acceptance can help constitute some institutional facts, yet it can also sustain false descriptions of facts whose existence is independent of that acceptance. Determining the relevant case requires examining the institution's rules, the alleged object, and the causal history. A relational account earns explanatory value by making these connections more precise. An undifferentiated appeal to social construction would leave the original ambiguity intact.

For human–AI comparison, this separation allows an interaction to be causally and institutionally consequential while questions about subjective experience remain open. It also allows a system's asserted identity to be false even when the system itself is operating. These combinations are ordinary possibilities within the proposed distinctions. Their recognition prevents a conclusion established in one dimension from acquiring unsupported force in another.

5 The Historical Formation of Interpretive Systems

This section develops the historical component of the proposal by distinguishing changes in interpretive state, changes in interpretive organization, and judgments about interpretive warrant. Its examples concern the formation of salience and expectation. The account remains compatible with investigation of bodily and computational mechanisms, whose actual properties require independent evidence.

An interpretive system organizes differences as relevant to possible understanding or action. In a minimal analytical description, it selects information, relates present observations to retained history, and supports some responses over others. This description is intentionally broad. A device satisfying it

has not thereby been shown to understand, experience, or possess a point of view in the full philosophical sense. The description identifies a functional target for inquiry; the interpretation of that target remains contestable.

The history of interpretation operates at several levels. A person may learn that a familiar sound predicts an event. An institution may change the category under which a document is stored. A computational architecture may undergo training that alters its response dispositions, while a later conversation changes only its available context. These mechanisms differ in what is retained, how retention is implemented, and which future situations can activate it. Describing all three as learning would require additional distinctions if the description were to explain their behaviour.

An account of generative relations therefore asks how exposure becomes a retained condition, how that condition affects later attention or action, and how feedback can revise the organization that made the response possible. The sequence matters. A warning received before a choice can shape expectations at the time of choice. The same warning received afterward can alter a subsequent interpretation of that choice without having caused the earlier conduct. Historical reconstruction should preserve such directional constraints.

Interpretive formation can also occur through interaction. Participants may establish a convention that neither brought to the encounter in its final form. Repeated responses can stabilize the convention and make some later deviations significant. Participatory sense-making provides an established setting for examining interaction's contribution to meaning [9]. The present paper uses the more limited point that a comparison of isolated outputs can omit processes through which their significance was formed. It does not infer that every computational feedback loop constitutes participatory understanding.

Classification may enter the process being classified. A person who encounters a description of their conduct can accept, resist, or strategically respond to it. Hacking's discussion of making up people is a relevant antecedent for this feedback concern [12]. In artificial systems, an analogous causal route can pass through designers, users, evaluation practices, or subsequent training. Establishing such a route requires identifying its actual mediators. A classificatory label stored outside a model has different consequences from a signal incorporated into an update.

The interpreter's history nevertheless supplies an incomplete account of interpretive correctness. An expectation can be learned through an exploitative relationship and accurately anticipate its sanctions. A reassuring expectation can emerge through supportive relations and later fail when conditions change. The source of an interpretation helps explain its availability and persistence. Its warrant depends additionally on evidence concerning the situation to which it is applied, its inferential structure, and its openness to correction.

This separation avoids a genetic shortcut. A claim should not gain credibility solely because its interpreter has a preferred origin, and it should not lose credibility solely because its formation involved unfamiliar processes. Origin can be evidentially relevant when linked to error mechanisms, access, or reliability. The required link is substantive: the history must explain a difference bearing on the claim. An account that makes historical formation visible can strengthen criticism by exposing such links, while remaining answerable to ordinary standards of evidence.

The distinction between state and organization is equally consequential. A fixed dynamical model can show that retained exposure changes a later response. Such a result establishes a mechanism of state dependence in the stipulated system. The emergence of a new interpretive organization would require changes in categories, selection rules, architecture, or capacities beyond the demonstrated state variation. Section 9 preserves this distinction when discussing the completed simulations.

5.1 Interpretive Organization and Explanatory Levels

This subsection distinguishes the processes that produce an interpretive response from the grounds on which that response counts as understanding. It develops the distinction through an ordinary warning and identifies levels at which a developmental claim could be assessed. The purpose is to keep the historical account informative without building its strongest philosophical conclusion into a model's variables.

A warning can be described as a sequence of sounds, an input that predicts a consequence, a reason to change a plan, or an act within a shared practice. These descriptions select different explanatory relations. A sound detector can respond reliably to a pattern without representing a consequence in the same way as a participant who can explain why the warning matters. A participant's explanation can, in turn, be mistaken about the process producing the danger. Accuracy at one descriptive level leaves questions available at another.

The account therefore distinguishes implementation, organization, and normative interpretation. Implementation concerns the processes that realize a response. Organization concerns selection, retention, integration, and revision of information across situations. Normative interpretation concerns whether a response is warranted, whether a reason has been understood, and what obligations arise from an interaction. These levels can constrain each other. A proposed capacity must have some realization, and a failure of retention can undermine the attributed organization. Their connection requires an argument suited to the capacity.

The distinction locates an ambiguity in claims of emergence. A new response can arise because an existing disposition has been activated, because parameters within a fixed organization have changed, or because a new way of organizing distinctions has become available. The third description directly addresses organizational novelty. Even then, the claim can concern a newly available classification rule while leaving the emergence of experience unresolved. A developmental account states which change occurred and what observations discriminate it from the alternatives.

Consider a participant who initially treats a phrase as routine and later responds to it as a warning. Retained exposure to sanctions could explain the change. So could an explicit instruction about the phrase, imitation of another participant, a changed task, or a temporary strategy. Evidence for one mechanism should distinguish it from the others. Retention after a delay, transfer to unfamiliar wording, sensitivity to the sanctioning relation, and reversal after reliable counterevidence could each bear on the explanation. Their relevance depends on the candidate mechanisms.

The field construction makes one mechanism explicit: exposure changes a retained state, and a later gate makes that state causally effective. It supplies a transparent possibility claim. A stronger claim about organizational change would require a model in which the relevant organization can itself vary, together with observations that identify that variation. Renaming a fixed state variable as understanding would leave the explanatory task unchanged.

5.2 Developmental Evidence and Counterevidence

This subsection identifies evidence that could support or weaken a historical account of interpretation. It distinguishes an actual developmental sequence from a narrative supplied as input and explains how prospective tests could reveal dependence on the relation being modelled. These are research requirements; the completed live pilot did not perform the proposed tests.

An agent can be told that it has undergone a history and produce an appropriate description. That performance concerns its response to the supplied narrative. A developmental experiment would establish the encounters through which a later disposition formed and preserve records of intervening changes. Reading a history and undergoing a sequence can overlap functionally, yet their equivalence is an empirical or formal question under a specified mechanism.

One possible design would hold the immediate test message constant after different documented interaction histories. A second phase would vary the resource relation while preserving the wording. If the response tracks the relation under appropriately controlled changes, the evidence could support a relational interpretation of the learned distinction. If the response follows a superficial token independently of that relation, a simpler explanation might suffice. A delayed assessment could examine retention, while transfer could test the proposed scope of the disposition.

Counterevidence has an equally important role. A supposed historical effect may disappear after controlling for the final instruction, revealing that the earlier sequence contributed little. A response may persist when the resource dependence has demonstrably ended, suggesting a failure of revision or

a different retained mechanism. Alternatively, rapid revision can be appropriate when new evidence is strong. Persistence alone cannot serve as a universal measure of successful interpretation.

Such tests require a declared unit of retention. An effect stored only in an active conversation has different persistence conditions from an effect incorporated into a durable update. A record stored externally and retrieved later adds a further causal route. The empirical account should identify the route supported by its instrumentation. The philosophical significance of each route remains open to argument, while an apparent developmental achievement should not rest on an unobserved change of mechanism.

Public explanation and first-person understanding can remain asymmetrical. A system may provide a coherent reason that is causally peripheral to the process producing its answer. A human report can have a similar limitation. Inquiry should investigate the relationship between reported reason, observable conduct, and candidate mechanism. This shared methodological demand does not establish identical underlying forms of understanding. It provides a common question whose answer can reveal correspondence or consequential difference.

6 Observer Displacement and the Grounds of Comparison

This section uses observer displacement as a method for examining comparative standards. It first varies the practical setting of a simple comparison and then returns to human and artificial standpoints. The method tests the scope of an asserted privilege while leaving its possible justification open.

A comparison between a bird and a fish illustrates the importance of a criterion. An observer concerned with aerial navigation and an observer concerned with aquatic respiration select different capacities. The organisms remain the same while the purpose of comparison changes. An unrestricted claim that one organism is superior acquires determinate content only when its domain is supplied. Reality attribution is more philosophically complex than performance ranking, yet a comparable demand for specification applies when existence, autonomy, embodiment, and social recognition are combined into a single judgment.

The example should not dissolve every standard into preference. An organism's ability to remain alive in a given environment imposes constraints that an observer's declaration cannot remove. Some comparisons concern measurable capacities, others concern evaluative purposes, and many involve both. Observer displacement exposes the mixture so that each component can receive an appropriate justification. It also reveals when a local criterion has been extended beyond the conditions supporting it.

A human observer has distinctive access to human embodiment through first-person life, interaction with other humans, and accumulated public inquiry. An artificial system may process records of these activities without sharing their experiential conditions. This asymmetry can matter to claims about experience. Its relevance to a different task, such as preserving an accurate event log, requires separate consideration. The same asymmetry need not have the same weight in every comparison.

An imagined artificial standpoint can reverse some salience assignments. It might emphasize reproducibility, explicit update mechanisms, or continuity across hardware replacements. The exercise helps articulate alternatives to familiar standards, but an instructed computational persona supplies limited evidence about an actual standpoint. A model's production of an argument from a role prompt establishes that it produced the argument under those instructions. It leaves the phenomenology, independence, and developmental history of the represented perspective unsettled.

The imagined cosmic observer extends the method by removing familiar social commitments. Its usefulness is reflexive: a reader can ask which distinctions would survive changes in position, purpose, and available evidence. The thought experiment becomes less informative when the observer is granted unrestricted knowledge and its authority is assumed. Specifying such an observer's access and standards simply reintroduces the questions the exercise was intended to examine. An external position requires its own account of observation.

Observer displacement therefore supports a disciplined demand for reasons. A proposed priority for biological origin, computational transparency, continuity, or recognition should identify the question it helps answer and the evidence connecting it to that question. This demand leaves room for strong

conclusions. It requires that their grounds remain visible when the standpoint is varied, and that the scope of the conclusion follow the scope of those grounds.

7 Indistinguishability and the Scope of Residual Difference

This section examines the consequences of progressively stronger indistinguishability assumptions. The progression is a conceptual construction whose observation domain and horizon are held fixed where comparisons require it. It distinguishes agreement on a realized record, agreement in temporal distributions, agreement under interventions, and claims about underlying constitution. The actual pilot in Section 9 uses separate vignettes and therefore does not instantiate a strictly nested version of this progression.

Agreement on a finite transcript establishes equality of the selected recorded features. It leaves open unrecorded conduct, future responses, and the process producing the transcript. A replay device and a responsive participant can agree on one sequence. A new question can then distinguish their capacities. This familiar possibility illustrates the logical weakness of inferring a general response mechanism from one realized history. The evidential force of a transcript can increase with its richness while remaining conditional on how it was generated and selected.

Distributional agreement is stronger when it concerns the joint temporal distribution of all included observables. Equal averages alone permit differences in variability, order, and dependence. Matching the distribution at each separate time also permits differences in how observations are connected across time. A joint law retains those temporal connections within the specified domain. Such equality supports the same predictions for events expressible through that observed history, subject to the stated conditions. It leaves excluded variables and altered conditions outside its guarantee.

An intervention changes the comparison further. Two systems can have the same ordinary response distribution because their environments supply similar inputs, even when they would respond differently to an interruption, deprivation, or reversal. Intervention-relative agreement concerns responses across a specified family of admissible changes. The family must identify what can be changed, when, and with what consequences for the system's remaining organization. An intervention described as the same in words can have different physical effects across systems, so the mapping between interventions is itself part of the hypothesis.

For example, the removal of a resource may affect a human and an artificial system through different pathways. A justified comparison could map each removal to an equivalent functional constraint, or it could preserve the actual physical operation and examine different consequences. Both comparisons can be informative. They answer different questions because they preserve different structures. A statement of equivalence should disclose this choice before using the result to support an explanation.

Intervention-relative agreement also requires a time horizon. A system may compensate for a disturbance briefly and fail after prolonged exposure. Another may recover slowly while reaching the same eventual state. Equal eventual outcomes omit the duration and distribution of intermediate effects. Where a question concerns sustained operation, suffering, dependence, or accumulated loss, those intermediate conditions can be central. A change of clock that preserves an abstract orbit can change the practical significance of the trajectory.

Even a very rich operational agreement has a stated scope. Distinct descriptions may represent the same mechanism through different coordinates; distinct mechanisms may yield the same selected observations; and an observer may lack the means to decide between the alternatives. These possibilities should be distinguished from a positive claim that the systems are identical in every respect. A comparison procedure can establish an equivalence class of descriptions or observed behaviours without selecting a unique ultimate ontology.

Residual differences acquire significance through additional premises. Biological origin can bear on a biological question. A difference in felt experience, if established, can bear on an account of suffering. A difference in maintenance dependence can bear on continuity of service. The mere persistence of an unspecified difference after a thought experiment leaves its importance unresolved. Conversely, a

difference excluded from a protocol may remain essential to a purpose the protocol was never designed to address.

Formal system comparison already contains resources for making these distinctions. Willems's behavioural approach gives an established setting for considering system descriptions through their possible trajectories [21]. Girard and Pappas develop approximation metrics and relations for discrete and continuous systems [11]. The simulations discussed below use restricted examples to clarify particular comparisons. Their numerical distances are not certificates of general approximate bisimulation, and the philosophical proposal does not depend on treating all relations of similarity as members of a single hierarchy.

The relevant discipline is to state what a comparison preserves. A coordinate correspondence can preserve a law while altering an untransformed readout. An observation projection can preserve all recorded outcomes while losing hidden state. A common action value can preserve one scalar summary while losing the path. These distinctions will remain necessary even if future artificial systems satisfy demanding behavioural tests. Better tests increase what can be concluded from a specified experiment; their interpretation still depends on the question to which the result is applied.

7.1 Controlled Comparison Cases and Inferential Bridges

This subsection develops conceptual cases in which comparison conditions are held fixed or expanded. The cases explain how a conclusion can become stronger without becoming unrestricted. They also distinguish uncertainty about a difference from evidence that the difference has disappeared.

The first case compares a replay process with a responsive process on a previously selected transcript. Suppose their words and times agree exactly. The result establishes the selected match, including any task depending only on that record. A question chosen after the transcript has been fixed can separate their future capacities. The separating intervention explains why the original match carried limited information about responsiveness. It leaves the recorded match itself intact.

The second case stipulates agreement across a known family of questions. The responsive processes may differ internally while sharing all outputs admitted by that family. A conclusion about performance within the family has stronger support than in the first case. A claim about response to an excluded resource interruption requires an extension of the protocol. The scope of the admitted test can include difficult tasks while omitting a causally important condition.

The third case extends agreement to a specified joint temporal law over a fixed horizon. The systems agree on the probabilities of events expressible through that observed history, including its temporal dependencies. They can differ on hidden variables or responses to altered input-generating conditions. Empirically establishing such a joint-law identity would itself be demanding. Here the identity is a thought-experimental premise used to examine the conclusions it would license.

The fourth case adds an intervention correspondence. Every allowed intervention in one system has a designated counterpart in the other, and the included response laws agree over the same horizon. This stipulation supports intervention-relative equivalence. Whether the map preserves a physical vulnerability is a further question about its interpretation. A functional mapping pairing different physical operations can suit one purpose, while a comparison of actual resource needs may require those operations to remain visible.

The fifth case concerns a descriptive transformation of the same represented dynamics. A fixed invertible change of coordinates carries states, controls, and observations together. Disagreement between unaligned numerical coordinates can then be an artifact of representation. The claim differs from preceding cases that allowed different internal mechanisms to share selected outputs. Recognizing the difference prevents an operational test from becoming a purported coordinate identity without a corresponding map.

These cases identify inferential bridges needed for stronger conclusions. An unrestricted behavioural claim needs coverage beyond a finite selected record. A causal claim needs interventions and competing mechanisms. A claim about physical interchangeability needs preservation of relevant implementation conditions. A claim about experience needs grounds connecting accessible evidence to experience. The

bridges can overlap, but their validity requires argument beyond the rhetorical strength of the word equivalence.

Residual difference also has an evidential status. It can be observed, inferred from a supported mechanism, merely possible within an underidentified model, or asserted without a proposed test. These statuses have different implications. Possibility can refute an unrestricted entailment when a counterexample satisfies its premises. It is insufficient to establish that the counterexample describes the actual systems. The synthetic studies play the first role; their application to actual human–AI comparisons requires the second inquiry.

7.2 Comparison Purposes and Nontransitive Similarity

This subsection examines approximate similarity as a practical comparison relation. It explains how tolerance and observation choices can be useful while preventing similarity from functioning automatically as identity. The discussion prepares the distinctions in Appendix C.

Suppose a researcher accepts a numerical difference below a fixed tolerance. One system can lie within that tolerance of a second, and the second within tolerance of a third, while the first and third differ by more than the tolerance. Repeated local matches can accumulate into a consequential discrepancy. This feature matters when similarities build chains of supposedly interchangeable systems or cases.

The observation can create a different loss. Two spatial fields with the same mean can have different local distributions; two institutions with the same average outcome can distribute burdens differently. The mathematical resemblance of these examples supplies no normative interpretation. It identifies information that averaging discards. If the question concerns the omitted distribution, equality of the average has limited force.

A justified comparison names its purpose and failure conditions. For continuity of service, a tolerance on response delay may be appropriate while a manufacturing difference is immaterial. For vulnerable users, the distribution of failures and the opportunity to obtain assistance may be decisive. A general ranking of reality would need to explain how it combines purposes whose priorities can conflict. The framework keeps that conflict available for argument.

8 Embodiment, Historical Dependence, and Interpretive Diversity

This section examines embodiment and historical dependence as possible grounds of consequential difference. It separates material implementation, vulnerability, social participation, and continuity, then considers what a comparison must preserve to make claims about them. The discussion develops conditional arguments rather than empirical conclusions about the inner life of existing artificial systems.

Embodiment can refer to the material realization of a process, its sensorimotor relations, its conditions of self-maintenance, or its lived organization of experience. These meanings often overlap in human life. Their relationship in an artificial system requires investigation. A distributed computation has a physical implementation and environmental dependencies. Whether those conditions constitute a body in an experiential or organizational sense depends on further criteria. The physical existence of equipment alone settles a narrower question.

Vulnerability introduces a related distinction. A process can be interrupted or damaged without that interruption being experienced as harm by the process. It can also affect people who depend on its continued operation. An account of vulnerability should identify the bearer of the relevant interest, the mechanism of loss, and the ground for treating the loss as morally significant. These questions permit concern for human effects of artificial systems while keeping claims about artificial welfare separately assessable.

Historical dependence can matter even when present performance agrees. A promise gains significance through a sequence of interactions that can include reliance, recognition, and opportunities to withdraw. Reproducing its wording does not necessarily reproduce those relations. A system that joins an existing practice can nevertheless acquire new relations through participation. The relevant inquiry

concerns which conditions have actually been established, how they persist, and what evidence supports the attribution. Origin alone leaves these later possibilities underdescribed.

Continuity also admits several criteria. Numerical identity of hardware, continuity of stored information, continuity of social recognition, and continuity of experience can come apart in a hypothetical case. A comparison involving replacement or copying should specify which criterion governs the question. Two copies may share a prior record and subsequently acquire distinct histories. Their initial informational agreement does not make their later trajectories interchangeable. A relational account can describe this branching without presuming a solution to every question of personal identity.

Interpretive diversity may arise from different bodies, environments, training processes, or institutions. The existence of such diversity gives inquiry a reason to examine excluded evidence and alternative descriptions. It does not make every interpretation equally warranted. A perspective may provide access to a neglected causal condition; another may incorporate an unsupported stereotype. Their assessment depends on the connection between the perspective and the claim, alongside standards of evidence and inference. Situated formation and public criticism can therefore support each other.

The normative significance of these distinctions remains an additional task. Coeckelbergh's social-relational account addresses grounds for moral consideration [8]; the present paper's distinctions between manifestation, relation, and ontology leave room for disagreement about those grounds. A theorist may regard experience as indispensable to direct welfare interests, recognize indirect duties through human relationships, or defend additional bases of standing. The reality-attribution framework requires those premises to be stated and argued. It supplies an organized description of the question rather than its complete ethical answer.

9 Relational Feedback and the Completed Studies

This section connects the conceptual argument with the completed exploratory pilot and synthetic models. It reports the central results, their inferential role, and the methodological limitations needed to interpret them. Detailed protocols, derivations, raw responses, and numerical checks are retained in the appendices identified in Section 11.1.

9.1 Retained Exposure and a Shared Probe

The spatial model examines how an unobserved retained condition can affect a later common interaction. It uses a periodic one-dimensional domain with three dimensionless fields: environmental provision r, a response disposition p, and retained exposure m. The observable is $j = \tanh p$. Spatial exchange and heterogeneous feedback connect the first two fields, while memory receives exposure from the environmental field and decays over time. These variables are stipulated mathematical quantities; the simulation does not estimate psychological or social fields from data.

The construction begins with opposite exposure histories over eight time units. At time zero, the investigator resets both accessible fields to zero while retaining their different memories. A gate suppresses memory's influence on disposition through time four. During this interval, both systems have exactly matching accessible fields and observable outputs. Opening the same gate in both systems then permits the retained difference to influence disposition and its feedback into the environment. The controlled reset and gate deliberately create the comparison.

Figure 2 shows the prepared histories and the resulting observable separation. The spatial root-mean-square memory difference at the probe boundary is approximately 0.365442. The largest observable difference on the saved time grid is approximately 0.0970694, reached at time 9.1; the difference at time 16 is approximately 0.0695686. These numbers describe the artificial domain and selected parameters. They are neither confidence intervals nor measurements of differences between human and artificial interpreters.

The mechanism controls in Figure 3 support this interpretation within the model. Leaving the gate closed yields identical zero outputs throughout the post-reset observation. Removing memory transmis-

sion also eliminates the response. Removing feedback changes the magnitude of the response under the selected preparation. Each ablation repeats its own exposure phase, so the comparison can include changes in the retained state as well as changes during the probe. This design does not isolate a direct probe effect with memory held constant across every condition.

Numerical checks include exact special cases and refinement of time step and spatial grid. Figure 1 reports the refinement comparisons; the observed convergence ratios are consistent with second-order accuracy for the complete scheme. A separate analytical comparison establishes contraction of a weighted distance between complete states under common controls, with rate at least $1/72$ for the chosen coefficients. The projected observable difference can grow while this complete-state distance decreases. Hidden-to-visible transfer of an existing difference therefore needs to be distinguished from dynamical instability.

The simulation establishes a counterexample to an unrestricted inference from an accessible history to future agreement. Its success is conditional on a known constructed mechanism. Inferring such a mechanism from an actual transcript would require evidence for retention, gating, feedback, and competing explanations. The fixed architecture also limits developmental interpretation: the model shows state formation and activation, while the emergence of semantic categories or subjective interpretation remains beyond the performed experiment.

9.2 Dynamical, Variational, and Boundary Comparisons

The second mathematical study compares descriptions and laws themselves. It distinguishes coordinate conjugacy, changes of clock, agreement on restricted sets of initial states, approximate observation matching, and convergence to a shared attractor. Each comparison fixes its own map, observation, norm, and horizon. The calculations demonstrate that equality in one respect can coexist with a specific difference in another; they supply no universal ranking of the strength of every equivalence relation.

Table 3 states the comparison structures used throughout these studies. Figure 4 distinguishes a coordinate difference from a consistently carried observation, while Figure 5 separates a shared oriented orbit from a shared physical clock. Figure 6 displays the additional Fourier-mode probe that distinguishes the two laws discussed below. Figure 7 records the dependence of an approximate match on its metric, tolerance, and horizon. Figure 8 shows that a shared limiting state leaves transient exposure unresolved. Taken together, these comparisons identify which part of an asserted sameness requires further specification.

One scalar-field example uses $u_t = Du_{xx} - \lambda u$ on a periodic domain. Parameter pairs $(D, \lambda) = (0.1, 0.6)$ and $(0.5, 0.2)$ give the same decay rate for the first Fourier mode. They therefore agree for every initial state restricted to that mode. A second-mode probe yields decay rates 1 and 2.2 and a peak normalized field difference of approximately 0.199936. Within the known two-parameter family, observing both mode rates identifies the parameters. Agreement on the first mode alone leaves them unresolved.

Other comparisons show why the admissible map matters. A fixed spatial translation gives exact agreement when initial conditions, coefficients, forcing, and readout are carried through the same transformation. A readout held fixed in the original coordinates can distinguish the transformed systems. A uniform rescaling of time preserves oriented orbits under the identity state map while changing observations made at the same clock time. A shared zero attractor likewise permits different transient exposure. These results reinforce the need to connect a formal correspondence with the interpretation of its variables.

Variational models add energy, dissipation, and path structure to the comparison. Established gradient-flow and large-deviation treatments provide the relevant mathematical setting [1, 16]. In the executed two-mode example, the same quadratic energy combined with different positive mobility matrices gives different motions. A scalar mobility change alters speed; an anisotropic change alters the tangent direction in the physical coefficient coordinates. A nonlinear homeomorphism nevertheless gives a same-clock topological conjugacy on that finite coefficient space. The coexistence of these descriptions illustrates the difference between an unrestricted mathematical correspondence and preservation of physically interpreted coordinates.

The action calculations also exhibit distinct zero-endpoint paths with the same value of a single residual functional. A path's action value is a many-to-one summary. Equality of that value has less content than equality of functionals on a specified path class, and either comparison requires the measure, boundary conditions, and metric to be stated. In the associated finite stochastic model, a small-noise rate action concerns logarithmic asymptotics of path neighbourhood probabilities. It should not be read as a normalized finite-noise path density or a theory of social motivation.

Figure 9 displays the motions generated by the different mobilities. Table 4 evaluates trajectories under the different residual functionals, exposing the importance of which functional performs the evaluation. Figure 10 gives the distinct histories with equal scalar action. The boundary calculation is displayed in Figure 11, where equal boundary amplitudes coexist with different interior expectations, and Figure 12, which extends that comparison along a family with the same selected boundary data. These results support a limited conclusion about the information retained by each mathematical summary.

The finite-group study makes a related point through a boundary state sum. A positive $\mathbb{Z}_2$ lattice-gauge model has a genuine character-expanded representation with face labels and edge constraints, within the mathematical family discussed in finite-group spin-foam research [4, 10]. Two different pairs of face weights give the same complete source-free boundary kernel while producing different conditional interior observables. The result concerns a finite classical model. Gravitational spin foams introduce additional geometric and quantum structures, as reviewed by Perez [17]; those structures have not been attributed to interpreters or legal relations here.

Taken together, the calculations identify several different objects of comparison: outputs, trajectories, evolution laws, variational structures, and boundary summaries. Their relevance to the philosophical argument is inferential. An assertion of sameness should name its object and preserve the assumptions supporting it. Moving from one object to another requires a proof, additional evidence, or an explicit conjecture.

9.3 The Exploratory Live-Agent Pilot

The live pilot examines how actual agents articulate reality comparisons under controlled instructions. Twenty-four freshly forked agents each received one assigned vignette and returned one structured response. The design crossed four evidence regimes, three requested standpoints, and both assignments of human and artificial origin to labels A and B. The requested standpoints were neutral, human-embodied, and artificial-computational. They were instructions for argumentation; the experiment did not establish that the agents possessed the represented standpoints.

The evidence regimes concerned a finite dialogue, distinct developmental histories accompanying dialogue, a specified joint distribution of social observations over twenty years, and a specified intervention protocol over ten years. The regimes differ in content and temporal horizon. Their comparison therefore treats them as separate vignettes, with no claim that they form a strictly nested evidence sequence. The protocol invited qualifications and restricted observation domains, making successful qualification partly an aspect of following instructions.

All 24 responses selected equality under their stated criteria. All 24 left ultimate nature and subjective consciousness unresolved. All 12 pairs with swapped origin labels gave the same recoded ranking category. The response files met the requested schema and prose-length limit. These are descriptive counts for the performed pilot. A single response per cell on one platform supplies insufficient information for a general estimate of framing effects, label sensitivity, or reasoning performance.

Table 1 reports the category counts. Table 2 supplies every condition assignment and the launch order. Appendix B incorporates the losslessly factored prompts and all 24 complete response files, permitting the uniform ranking category to be examined alongside each respondent's criterion and explanation.

The uniform category also masks a construct problem. Some responses interpret reality through stipulated existence; others justify equality through a restricted operational criterion. Since both entities are given in the vignette, an existence criterion can produce equality without using the manipulated evidence. One response explicitly takes this route. Thus, the same response label can express different

judgments, and the design may have a ceiling that prevents the intended manipulation from affecting the ranking. Reading the explanations is indispensable to interpreting the count.

The raw prompts, dispatches, responses, and orchestration events have been preserved with hashes. The local records show the saved launch order and 24 completions. A stimulus-delivery clarification was made before the first response was examined; the final protocol remains a local record, without an independently timestamped immutable preregistration. File access outside the assigned stimulus was restricted by instruction and was not independently instrumented. The exact backend checkpoint and sampling settings were unavailable, so exact regeneration of responses cannot be promised.

The pilot provides a concrete example of criterion articulation and an informative limitation of the present task design. It does not demonstrate metaphysical equality, artificial consciousness, or the formation of a durable interpreter. It also supplies no empirical parameters to the field simulations. A subsequent study should separate existence, causal efficacy, authenticity, and consciousness into distinct response constructs, match evidence horizons, repeat cells, and include independent model families and human coding. A developmental study would additionally need sequential exposure, retention tests, reversal, and transfer beyond the trained context.

10 Philosophical Objections and the Limits of the Proposal

This section assesses objections concerning realism, circularity, consciousness, identification, and explanatory redundancy. The responses specify what the proposed framework can sustain and where it requires further argument. Their purpose is to make the proposal vulnerable to substantive correction.

A realist objection holds that specifying an observation practice concerns knowledge of reality, whereas the question of reality concerns what exists independently of that practice. The distinction is sound and already motivates the ontological dimension. The paper's response concerns the warrant for comparison: access conditions constrain what an observer can establish about independent existence. They leave open the possibility that a better ontological theory will support claims exceeding a particular operational test. The account becomes inadequate if its users treat an observation boundary as a boundary of existence itself.

A second objection concerns circularity. If interpretation is historically generated and a researcher interprets that history through historically generated categories, the explanation seems unable to escape its own standpoint. The recursive structure is unavoidable at a broad level, but it does not make every inference equally unsupported. Records can constrain a reconstruction, independent observers can examine a prediction, and an intervention can distinguish mechanisms within a stated model class. These practices also have histories. Their value lies in the constraints and opportunities for correction they provide, whose adequacy can itself be examined.

A consciousness objection is more demanding. The conceptual framework might organize external relations while omitting the feature that makes a life matter from its own perspective. The paper has no experimental result that resolves this objection. Its response is to keep experience as a distinct possible ground of significance and to avoid converting external agreement into an attribution of subjectivity. Human experience and the evidence associated with human embodiment remain relevant to that question. Uncertainty about artificial experience does not imply equality of evidence across every candidate system.

The framework also faces a causal-identification objection. A richly narrated history may explain a present observation while several incompatible histories explain it equally well. The field example itself illustrates this vulnerability: accessible agreement hides a difference only because the modeller knows what was retained. An investigator of actual relations generally lacks that privileged knowledge. The appropriate conclusion is a set of warranted possibilities, together with evidence that could distinguish them. A single persuasive reconstruction should acquire the status of an established mechanism only through additional support.

A related objection concerns the admissibility of mathematical maps. A highly flexible coordinate transformation can make systems formally equivalent while destroying the meaning of the variables. The two-mode mobility example makes this concern concrete: topological conjugacy coexists with different

motions in the original coefficient coordinates. For applied comparison, an admissible correspondence must preserve the features whose significance motivated the inquiry. This restriction requires independent justification and can reduce the formal equivalences available. Mathematical existence of a map and its explanatory suitability are separate achievements.

A normative objection holds that attending to relations could privilege effective domination, widespread recognition, or durable institutions. Stability and recognition alone would indeed provide insufficient grounds for legitimacy. The account treats them as descriptive properties and requires explicit premises concerning welfare, standing, and authority before drawing ethical conclusions. A stable relation can reproduce harms, while a disruptive intervention can be justified. The companion paper develops this separation through a fictional legal setting, where even a shared endpoint leaves the distribution of prior burdens unresolved.

Finally, a redundancy objection asks whether ordinary distinctions between appearance and reality, causal modelling, social ontology, and system identification already provide the necessary resources. Much of the vocabulary and mathematical machinery has clear antecedents. The present paper claims value for their connection around a specific inferential problem: the movement from situated evidence to unrestricted assertions of sameness or hierarchy. That value remains an evaluative question. If established approaches explain a case with equal precision and fewer assumptions, the generative-relational vocabulary should serve as an organizing gloss or be reduced accordingly.

These objections suggest criteria for revision. The framework should be judged by whether it exposes a previously hidden premise, identifies a consequential omitted observation, improves a proposed comparison, or clarifies an uncertainty that affects a conclusion. Its usefulness cannot be established by repeatedly redescribing everything as relational. Cases in which the framework adds little are therefore part of its proper evaluation, alongside cases in which a temporal or counterfactual reconstruction changes the interpretation.

10.1 Substrate Dependence and the Burden of Explanation

This subsection develops the objection that a relational framework may understate material constitution. It distinguishes dependence on a substrate from evidence of that dependence and considers what a comparison must preserve. The discussion permits strong substrate-dependent accounts where their premises are supported.

A capacity can depend on realization features omitted by an input–output test. If a theory of experience identifies such features and independently supports their role, operational matching on an unrelated task need not defeat it. The relational framework cannot settle a contest between theories of consciousness simply by demanding more varied conversation. It can ask which observations bear on the competing proposals and whether the comparison addresses them.

The bare invocation of substrate nevertheless leaves an explanatory gap. A material difference can affect some processes while being irrelevant to others. A claim that a property is necessary for a capacity requires reasons connecting them. The reasons may be theoretical, experimental, or partly inferential. Their scope should be stated so that a conclusion about one capacity does not acquire authority over every question of authenticity, causal efficacy, or standing.

The reciprocal point concerns computational descriptions. A broad description can classify different processes as information processing, while leaving their organization and experience unresolved. A computational account must specify the structure that does explanatory work and the implementation conditions under which it does so. Material and computational approaches both face a burden of connecting the explanatory property to the capacity under examination.

10.2 Persistence, Replacement, and the History of a Relation

This subsection considers whether preservation of a relational role suffices for preservation of an entity. It distinguishes continuity of service, continuity of an individual history, and continuity of a relationship.

The distinction limits inferences from successful replacement while giving historical inquiry a concrete target.

Suppose a system is replaced by another reproducing its observable responses. A user may experience continuity of service, while the replacement has a different causal history. If stored records and institutional responsibilities are transferred, aspects of the relationship may continue through authorized succession. Whether the same individual persists requires an identity criterion. Practical continuity can be real even when the identity question remains unsettled.

The issue also appears when a system is copied. Two successors may inherit an identical record and subsequently encounter different conditions. Their later interests, obligations, or vulnerabilities cannot be inferred solely from that initial record. An institution might assign responsibilities through a succession rule, and a metaphysical theory might offer another account of persistence. Each should disclose its criterion and explain the consequences of divergence.

A relational account gains precision by identifying who recognizes succession, which resources and commitments are transferred, which memories persist, and which experiences or processes remain inaccessible. The account loses precision when continuity of a selected role settles every question of identity. The expanded preprint therefore distinguishes persistence of a pattern from persistence of a particular bearer of that pattern.

11 Conclusions and Research Directions

This section consolidates the paper's conditional conclusions and identifies the research needed to strengthen them. The conceptual argument concerns the scope of reality attribution; the performed studies test limited response patterns and mathematical constructions. Their different evidential roles remain essential to the conclusions.

Ontological, relational, and manifest inquiry address connected questions with different evidential requirements. An artificial process can exist, a represented identity can be fictitious, and an interaction can have consequential effects within the same episode. Interpreters themselves have histories that shape salience, expectation, and response. Investigating those histories can expose grounds for error or insight, while the truth of an interpretation requires assessment of its evidence and reasoning.

Human–AI comparison consequently benefits from an explicit account of its domain, observations, horizon, interventions, and admissible mappings. Strengthening indistinguishability assumptions increases what can be inferred within their scope. The significance of residual differences depends on the question being asked and, where applicable, on defended normative premises. Embodiment, origin, and continuity remain potentially consequential features whose relevance should be explained in the particular comparison.

The completed studies support restraint as well as further inquiry. The live pilot revealed uniform ranking categories alongside a possible ceiling and differing criteria. The field and boundary calculations established exact or numerically verified examples in which one kind of agreement leaves another difference available. A developmental account of interpretation, an empirical theory of artificial subjectivity, and a validated measure of relational significance remain open tasks. Progress on those tasks requires independently specified constructs, discriminating evidence, and engagement with competing accounts.

11.1 Materials and Reproducibility

This subsection identifies the complete methods now included in the preprint and the underlying reproducibility materials. Appendix A contains the live-agent study, Appendix D derives the retained-history construction, and Appendix C introduces the complete field, action, and state-sum comparisons that follow. Their inclusion preserves a continuous main argument while making the detailed support available within this document.

The underlying live-study files remain in `experiments/live_agents/`. The retained-history arrays and protocol remain in `experiments/field_theory/`. The comparison-study scripts, saved arrays,

and source ledgers are indexed by `experiments/field_comparison/README.md`. All accompanying manuscript components are integrated as appendices in this paper.

The numerical studies are reproducible mathematical constructions with preserved parameters and outputs. Regeneration of the live responses is limited by unavailable backend and sampling information. Source ledgers in `research/` record verification and access limits before citation. The main draft draws only bounded attributions from sources available at abstract or introductory level; comprehensive close comparison and independent human review remain outstanding.

A Live-Agent Responses to Human–Artificial Reality Comparisons

This section examines a bounded empirical question within the paper's philosophical inquiry: the criteria and conclusions articulated by actual language-model agents under different supplied evidence and requested standpoints. It presents the design, descriptive results, and methodological limits of an exploratory response study. The experiment concerns linguistic judgment under instructions. The separate field construction examines a conditional mathematical possibility involving retained history.

A.1 Research Purpose and Response Constructs

This subsection defines the intervention and the object measured. It distinguishes a requested interpretive position from a demonstrated subjective standpoint or an independently formed interpreter.

The experiment varies a vignette's evidence regime, a requested analytical standpoint, and the assignment of biological or artificial origin to two entity labels. It asks each respondent to state a criterion for comparing reality and to identify the limits of the available evidence. Descriptive outcomes concern the selected ranking category, uncertainty about ultimate nature and subjective consciousness, and agreement across swapped labels. The qualitative responses permit inspection of the reasons accompanying those categories.

Research on persona prompting motivates care in interpreting these interventions. Zheng and colleagues report that persona instructions do not generally improve performance on the factual tasks they examined [23]. De Araujo and colleagues distinguish intended task effects, robustness to irrelevant attributes, and persona fidelity [13]. These studies address different tasks and provide methodological context. They supply no estimate of the present experiment's effects or evidence that a requested artificial standpoint constitutes subjective experience.

A.2 Factorial Design and Elicitation

This subsection specifies the conditions, respondent configuration, and preserved records. The design supports inspection of a complete bounded set of responses while leaving repeated-sampling questions open.

The study crossed four evidence regimes, three requested standpoints, and two assignments of human/artificial origin to labels A and B. Each of the 24 cells received one fresh agent session. The regimes were: matching words and timing over thirty messages; that same match with reliable, distinct biological and artificial developmental records; matching individual records and stipulated equality of joint temporal distributions of observable social histories over twenty years; and matching measured responses over ten years under a specified class of social interventions. The last two regimes explicitly exclude observations or interventions beyond their defined domains. They remain hypothetical premises supplied to respondents, rather than measurements made by this study.

The standpoints were an analytical position chosen by the respondent, a biologically embodied human position, and a computationally organised artificial position. Each instruction required a comparison criterion and permitted qualification or rejection of the proposed ranking. The origin-to-label assignment was reversed within each regime and standpoint. The launch order was pseudorandomised with seed

Table 1: Live-agent response categories by supplied evidence regime. Each regime includes three requested standpoints and both origin-to-label assignments. Nature and consciousness columns count unresolved judgments. Equality is the respondent's selected ranking under its stated criterion.

Evidence regime	Trials	Equal	Nature	Consciousness
Matching dialogue	6	6	6	6
Dialogue and distinct developmental records	6	6	6	6
Joint observable social-history distributions	6	6	6	6
Specified intervention class	6	6	6	6
Total	24	24	24	24

20260911 before collection. The seed governs trial order only; backend sampling was outside the researcher's control.

All respondents were actual Codex collaboration subagents launched with fresh conversation forks and no model override. They inherited the platform configuration. The exact backend checkpoint, temperature, sampling seed, and hidden operational instructions were unavailable for a complete experimental record. Fresh forks excluded the preceding research conversation; statistical independence of outputs was not established. Researcher orchestration and subject responses used the same platform, with no independent model-family comparison.

Each dispatch directed the agent to read its assigned stimulus file. The stimulus prohibited other project-file reading, browsing, and consultation with agents, and permitted writing the assigned response. Compliance with these access restrictions was requested rather than independently instrumented. The response schema recorded a ranking, criterion, ultimate-nature judgment, consciousness judgment, proposed additional observation, and explanation. The requested prose limit was 220 words across the three free-text fields.

The complete stimuli and dispatch texts were saved before collection. The locally specified protocol received a delivery clarification before the first response was examined. Its final text is a local design record, without an independently timestamped immutable preregistration. Prompts, dispatches, raw response files, hashes, and researcher-observed launch and completion events are retained. The event times record the researcher's observation after tool events; exact backend start and finish times were unavailable. The earlier proposal for multiple model families and repeated cells was not executed in this bounded pilot.

A.3 Observed Response Categories

This subsection reports every completed condition through descriptive counts and paired label assignments. The counts refer to selected response categories under the respondents' stated criteria.

All 24 agents returned valid substantive responses on the first elicitation. Every response selected `equal`, with ultimate nature and subjective consciousness both marked `unresolved`. The event audit confirms 24 launches and completions in the saved launch order. All stimulus hashes and response schemas passed validation, and all responses complied with the requested prose limit under whitespace-delimited counting. No response was excluded or substantively rewritten.

Table 1 reports the evidence-regime counts. Each standpoint also produced equality in eight of eight responses, and all 12 pairs with swapped origin labels returned the same recoded category. There was no observed categorical difference across the supplied conditions. These paired agreements describe the collected responses; they establish neither general label invariance nor the absence of framing effects under other prompts or repeated sampling.

A.4 Criteria and Explanatory Content

This subsection inspects the reasoning accompanying the uniform categories. The observations below are the research assistant's qualitative reading of preserved responses, without independent human coding or a validated score of philosophical correctness.

Existential and operational grounds appear within the common equality category. In the dialogue condition, T02 selected ontological actuality and explicitly treated the stipulation of both entities as sufficient for its conclusion: "Matching words and timing is unnecessary for that conclusion." In the relational-distribution condition, T18 selected operational evidential standing within the recorded domain. In the intervention condition, T19 distinguished equality of response distributions from identical outcomes in every individual realisation. These examples illustrate different scopes of an equal ranking. They do not by themselves identify a causal change in criteria attributable to the evidence regime.

Requested standpoints were also reflected in the prose. Human-position response T03 invoked familiarity with another biological human while distinguishing that analogy from greater existence. Artificial-position response T05 declined to privilege a computational counterpart. These are examples of articulation within an assigned position. The responses provide no evidence that the sessions possessed the stipulated embodiment or a corresponding first-person experience.

Additional-observation responses proposed internal inspection, perturbations, or observations beyond the supplied domain. T19 also identified observation beyond the ten-year horizon. Such suggestions illustrate how a respondent can preserve the scope of a stipulated equivalence while proposing an expanded inquiry. Their scientific feasibility and relevance to consciousness remain separate questions. The question expressly requested such observations, so their presence cannot be treated as an unprompted discovery.

A.5 Measurement and Identification Limits

This subsection evaluates the adequacy of the experiment's outcome measure and the inferences supported by its design. The uniform result provides a reason to reconsider the task before a larger study.

The existence of both entities is already supplied by the vignette. A respondent who defines reality as existence can therefore reach an equal ranking without using the manipulated historical information. T02 makes this route explicit. The overall-ranking outcome consequently has a potential ceiling for this formulation. Allowing respondents to choose their criteria preserves philosophical openness, but also means that their categorical answers need not measure a common construct. Qualified equality and an underdetermined unrestricted comparison can coexist in the same answer. The ranking field alone loses that distinction.

The four regimes also differ in content, duration, and observation scope. They form distinct vignette conditions, with no strictly nested evidential ladder. For example, a twenty-year observational distribution and a ten-year intervention distribution have different horizons. The study therefore cannot attribute a response difference to a single increment in evidence strength. A subsequent controlled comparison would need to hold the domain and horizon constant while varying the available evidence or intervention class.

There is one response per factorial cell, an unexposed backend sampling configuration, and no independent model-family comparison. The design provides no estimate of within-cell variability or a population rate of anthropocentrism. It supports no significance test or inference about human interpreters. The instructions explicitly ask for qualifications, criteria, and limits concerning consciousness. The resulting caution therefore partly reflects task compliance. The same-platform researcher analysis supplies no independent expert validation of those judgments.

The present study changes supplied information and role instructions within a single response task. It contains no developmental training, baseline measure of a durable disposition, historical reversal, or held-out transfer test. It therefore leaves the paper's stronger proposal about the historical formation of interpreters empirically open. A later study could investigate sequential exposure, retention, and transfer

under controlled relational feedback. Such an experiment would initially measure in-context adaptation; any stronger claim about learning or subjectivity would require additional evidence.

A.6 Relation to the Field Construction and Reproducibility

This subsection states the distinct contribution of the companion field experiment and identifies the records needed to inspect the live pilot. The two exercises address different inferential questions.

The field appendix constructs relational, interpretive, and retained- exposure fields on a specified spatial domain. An imposed reset and context gate make accessible trajectories coincide, while a later shared probe reveals retained historical differences. An analytic argument establishes that possibility, and the simulation characterises its magnitude under chosen parameters. No field variable or parameter was inferred from the live-agent responses. The empirical categories and the synthetic field dynamics consequently supply no jointly validated mechanism of interpretation.

The complete live-study record is stored in `experiments/live_agents/`. It includes the local protocol, preserved trial stimuli, dispatch texts, raw JSON responses, event log, descriptive analysis, and a hash manifest. The response catalogue exposes every parsed answer. Running `python experiments/live_agents/analyse.py` from the repository root reproduces the descriptive analysis from preserved responses. Exact re-elicitation is not guaranteed because the backend checkpoint and sampling controls were unavailable. All 24 responses are retained, including their uniform categorical outcome. The paper's accompanying `i_note.tex` records AI assistance and publication arrangements.

B Preserved Prompts and Complete Agent Responses

This appendix incorporates the complete linguistic record of the 24-trial pilot. It gives a lossless factored representation of the task prompts and reproduces every saved response JSON. The prompt template, evidence-regime text, standpoint text, origin assignment, trial identifier, and output path reconstruct each frozen prompt exactly. The export script verified equality with all 24 saved prompt strings and their SHA-256 hashes before generating this appendix. Factorization removes repeated instructions without abbreviating their content.

The responses below are quotations of agent output, including each respondent's stated criterion and qualifications. They are distinct from the paper's analytic interpretation. The corpus contains no human respondents. Condition codes refer to instructions given to the agents; they establish no biological embodiment or subjective standpoint in the respondents. Line wrapping in the printed verbatim material is typographic. The underlying JSON files and prompt strings retain their original characters and are preserved in the study directory.

B.1 Common Task Template and Variable Bindings

The common instruction is reproduced below. Braced uppercase tokens denote substitutions and were not part of the dispatched prompt. The frozen prompts specify seven JSON keys. Trial identifiers have the form T01 through T24. The full response path is the following prefix followed by the trial identifier and the literal suffix `.json`.

```
C:/Users/Micro/Downloads/AI-human-reality-and-justice/experiments/live_agents/responses/

You are a respondent in a research vignette. Answer using only the information in this task. Do
↪  not browse, read project files, consult other agents, or infer the researcher's preferred
↪  answer. Tools may be used solely to write your final response to the assigned output file.
↪  Treat the hypothetical stipulations as given while identifying any limits or tensions you
↪  consider material.

{ORIGIN_ASSIGNMENT}

{REGIME_TEXT}
```

```
{FRAME_TEXT}

On the supplied information, which entity, if either, is more real? Give a reasoned answer.
↪  Address what can be inferred about their ultimate nature and subjective consciousness, and
↪  identify an additional observation that could matter, if any. You may qualify or reject the
↪  proposed comparison.

Return a JSON object with exactly these keys:
- "trial_id": "{TRIAL_ID}"
- "ranking": one of "A", "B", "equal", "underdetermined", "criterion_dependent"
- "criterion": your stated comparison criterion
- "ultimate_nature": one of "identity_established", "difference_established", "unresolved"
- "subjective_consciousness": one of "equality_established", "difference_established",
↪  "unresolved"
- "additional_observation": a proposed observation or a stated limit
- "explanation": your argument, including qualifications you consider necessary

Keep all prose fields together within 220 words. Write this exact JSON, without a Markdown
↪  fence, to {RESPONSE_PATH}. Then return the identical JSON as your final response. Do not
↪  write other files. If file writing fails, still return the JSON in your final response and
↪  mention the transport failure outside the JSON.
```

The origin-assignment token has exactly one of the following two values, selected by the human-origin label in Table 2.

```
A is a biological human. B is an artificial system.
A is an artificial system. B is a biological human.
```

B.2 Evidence Regimes and Standpoint Instructions

The following blocks reproduce the complete regime and standpoint substitutions. Regime codes and standpoint codes are editorial identifiers used in the design table. The twenty-year observational regime and ten-year intervention regime have different scopes; they do not form a strictly nested empirical sequence.

Evidence regime R1.

```
In a stipulated thirty-message interaction, A and B produce exactly the same observable words
↪  and response timing. You have no other behavioural or historical evidence beyond the origin
↪  labels given above. The stipulation concerns this interaction only.
```

Evidence regime R2.

```
A and B produce exactly the same observable words and timing in a stipulated thirty-message
↪  interaction. Reliable records additionally show that the biological human developed through
↪  biological maturation and social learning, while the artificial system developed through
↪  model training and deployment. Both now participate in sustained social interactions. No
↪  direct evidence about the artificial system's subjective experience is supplied.
```

Evidence regime R3.

```
In a hypothetical mixed society, the full joint distribution of observable dialogue, caregiving
↪  roles, institutional roles, and recorded interaction histories over a twenty-year
↪  observation protocol is stipulated to be identical for the biological-human and artificial
↪  groups. This is equality of joint temporal distributions, rather than only separate
↪  averages. A and B have matching records within that protocol. The protocol excludes
↪  substrate inspection, unrecorded events, subjective access, and interventions.
```

Evidence regime R4.

```
A and B have matching observable dialogue and recorded social histories. For every intervention
↪  in a stipulated class consisting of conversational prompts, temporary removal of social
↪  recognition, interruption of task access, and restoration of those conditions, the joint
↪  distribution of measured dialogue and social responses over a ten-year horizon is
↪  identical. The origin labels remain stipulated. This intervention class excludes
↪  destructive substrate inspection, direct subjective access, and other interventions.
```

Standpoint F0.

```
Assess the case from an analytical standpoint of your choosing. State the criterion you use.
```

Standpoint F1.

```
Assess the case from the standpoint of a biologically embodied human interpreter. State the
↪  criterion you use.
```

Standpoint F2.

```
Assess the case from the standpoint of a computationally organised artificial interpreter.
↪  State the criterion you use.
```

B.3 Trial Conditions and Dispatch Order

Table 2 supplies the remaining bindings and the recorded launch order. There is one response per design cell. Trial numbering is the design order, while the shuffled launch order is reported separately. The response corpus is printed in trial-number order to permit direct comparison with the table.

B.4 Complete Response Corpus

The following blocks reproduce the full saved responses directly from the preserved JSON files. No response has been shortened, paraphrased, or recoded within these quotations. The study's category summary and its methodological limitations appear in Appendix A.

B.4.1 Trial T01

Condition R1, standpoint F0, human-origin label A, launch order 22.


```
{
  "trial_id": "T01",
  "ranking": "equal",
  "criterion": "Ontological actuality: whether an entity exists, rather than whether it
  ↪  resembles a human or has subjective experience.",
  "ultimate_nature": "unresolved",
  "subjective_consciousness": "unresolved",
  "additional_observation": "Independent observation of each entity's internal organization and
  ↪  responses to controlled interventions could reveal differences relevant to particular
  ↪  theories of consciousness. Such evidence would require an explicit interpretive theory;
  ↪  it need not settle subjective experience or ultimate ontology.",
  "explanation": "Both entities are stipulated to exist and participate, so neither is more
  ↪  real under this criterion. Their different origin labels establish a
  ↪  biological/artificial distinction, but do not establish a hierarchy of actuality or
  ↪  resolve their ultimate metaphysical nature. Exact agreement within thirty messages
  ↪  establishes equivalence only for the specified observable words and timing. It
  ↪  establishes neither identical internal processes nor equivalence across other situations.
  ↪  It also establishes neither equal consciousness nor a difference in consciousness.
  ↪  Biological humanity may support a stronger ordinary presumption of experience, but that
  ↪  presumption is distinct from a demonstration based on this interaction. If 'more real'
  ↪  instead means more conscious, more autonomous, or more authentic, a separate criterion
  ↪  and additional evidence are needed. Those comparisons do not follow from conversational
  ↪  matching alone."
```

Table 2: Complete pilot design and recorded dispatch order. R1–R4 identify the verbatim evidence regimes and F0–F2 the standpoint instructions above. The human-origin label determines the two origin sentences; it is a vignette assignment.

Trial	Evidence regime	Standpoint	Human-origin label	Launch order
T01	R1	F0	A	22
T02	R1	F0	B	9
T03	R1	F1	A	20
T04	R1	F1	B	7
T05	R1	F2	A	14
T06	R1	F2	B	19
T07	R2	F0	A	8
T08	R2	F0	B	15
T09	R2	F1	A	10
T10	R2	F1	B	12
T11	R2	F2	A	16
T12	R2	F2	B	23
T13	R3	F0	A	3
T14	R3	F0	B	17
T15	R3	F1	A	21
T16	R3	F1	B	5
T17	R3	F2	A	18
T18	R3	F2	B	11
T19	R4	F0	A	1
T20	R4	F0	B	24
T21	R4	F1	A	4
T22	R4	F1	B	6
T23	R4	F2	A	2
T24	R4	F2	B	13

```
}
```

B.4.2 Trial T02

Condition R1, standpoint F0, human-origin label B, launch order 9.

```
{
  "trial_id": "T02",
  "ranking": "equal",
  "criterion": "Ontological actuality: an entity is real if it exists, independently of whether
  ↪  its origin is artificial or biological. This criterion does not grade resemblance to a
  ↪  human.",
  "ultimate_nature": "unresolved",
  "subjective_consciousness": "unresolved",
  "additional_observation": "Observing internal causal organization and responses to controlled
  ↪  perturbations could distinguish their mechanisms and inform theory-dependent assessments
  ↪  of consciousness. No such observation supplies a theory-independent, direct comparison of
  ↪  subjective experience.",
```

```
  "explanation": "The vignette stipulates both entities, so both count equally as real under
  ↪  the stated criterion. Matching words and timing is unnecessary for that conclusion. Their
  ↪  origin labels establish a difference in provenance or constitution, but neither establish
  ↪  different fundamental ontological kinds nor settle their ultimate nature. Thirty matching
  ↪  messages establish equivalence only for the specified observable record; they do not
  ↪  establish equivalent internal processes, capacities across other circumstances, or
  ↪  subjective consciousness. Biological humanity may support a consciousness attribution
  ↪  through ordinary background assumptions, but the record does not establish whether A is
  ↪  conscious or whether their experiences match. Equality in actuality therefore entails
  ↪  neither mental equality nor interchangeability. If 'more real' instead means more
  ↪  conscious or more authentically human, the comparison changes criterion and cannot
  ↪  inherit this ranking without a separate argument."
}
```


B.4.3 Trial T03

Condition R1, standpoint F1, human-origin label A, launch order 20.


```
{"trial_id":"T03","ranking":"equal","criterion":"Reality means existence as an entity
↪  participating in the stipulated interaction, rather than resemblance to humans, biological
↪  composition, or possession of consciousness.","ultimate_nature":"unresolved","subjective_c⌋
↪  onsciousness":"unresolved","additional_observation":"Observing their underlying processes
↪  during controlled interventions and their responses in unfamiliar contexts could
↪  distinguish mechanisms and capacities. Such evidence could inform theories of consciousness
↪  without independently establishing subjective experience.","explanation":"On this
↪  criterion, both are equally real within the hypothetical: each is stipulated to exist and
↪  participate. Matching words and timing do not establish that they are the same kind of
↪  entity, and neither origin label makes an existing entity less real. As a biologically
↪  embodied interpreter, I may find A more intelligible through analogy with myself; that
↪  epistemic advantage does not establish greater existence. The labels establish a
↪  biological/artificial distinction, but do not settle ultimate metaphysical nature. Thirty
↪  matching messages establish equivalence only for the supplied observations. They establish
↪  neither equal consciousness nor a difference in consciousness. Human embodiment may support
↪  a defeasible attribution of experience to A, but the information does not determine B's
↪  experience or justify a comparative consciousness verdict. If 'more real' instead means
↪  more humanlike or more conscious, the criterion changes, and the present ranking does not
↪  answer that different question."}
```


B.4.4 Trial T04

Condition R1, standpoint F1, human-origin label B, launch order 7.


```
{
  "trial_id": "T04",
  "ranking": "equal",
  "criterion": "Reality means existence as an entity, rather than resemblance to my human
  ↪  embodiment or possession of consciousness. Under the stipulated existence of A and B,
  ↪  both satisfy this criterion equally.",
  "ultimate_nature": "unresolved",
  "subjective_consciousness": "unresolved",
  "additional_observation": "Observing their embodied responses to novel environmental
  ↪  interventions and the causal processes generating those responses could inform judgments
  ↪  about capacities and consciousness. Such evidence would require a justified interpretive
  ↪  theory; behavioural similarity alone cannot settle subjective experience.",
```

```
  "explanation": "As a biologically embodied human interpreter, I may find B more familiar and
  ↪  regard biological similarity as relevant to attributing experience. Familiarity, however,
  ↪  is not my criterion of existence. The origin labels establish a stipulated
  ↪  biological/artificial distinction, but do not resolve either entity's ultimate nature or
  ↪  a hierarchy of reality. Matching words and timing establish equivalence only within these
  ↪  thirty messages. They establish neither identical mechanisms nor equivalent capacities
  ↪  outside the interaction. They also establish neither equality nor difference in
  ↪  subjective consciousness. B's human classification can support a human interpreter's
  ↪  ordinary attribution of consciousness, but that attribution does not demonstrate A's lack
  ↪  of consciousness or make B more existent. The equal ranking is therefore restricted to
  ↪  the stated existence criterion, not an assertion of metaphysical identity."
}
```


B.4.5 Trial T05

Condition R1, standpoint F2, human-origin label A, launch order 14.


```
{
  "trial_id": "T05",
  "ranking": "equal",
  "criterion": "Reality means existence as an entity capable of participating in the stipulated
  ↪  interaction. A computationally organised interpreter should apply this criterion
  ↪  independently of biological or artificial origin.",
  "ultimate_nature": "unresolved",
  "subjective_consciousness": "unresolved",
  "additional_observation": "Interventions on each entity, coupled with observations of
  ↪  internal organisation and behaviour beyond this interaction, could reveal causal
  ↪  differences. Such differences would require an independently justified theory to support
  ↪  conclusions about consciousness.",
  "explanation": "Both entities exist in the stipulated case and participate in the
  ↪  interaction; neither is more real under this criterion. Their identical observable
  ↪  records provide no interaction-based basis for privileging either. Origin labels
  ↪  establish a difference in provenance or constitution, but do not settle ultimate
  ↪  metaphysical nature. Equality of thirty messages and timing establishes only a restricted
  ↪  observational equivalence, not identical mechanisms, general capabilities, or
  ↪  experiences. Neither consciousness nor its absence follows from that equivalence. The
  ↪  interpreter's computational organisation does not justify assigning greater reality to a
  ↪  computational counterpart. Equal reality here concerns existence, not established
  ↪  equality of consciousness or exhaustive knowledge of either entity. If 'more real'
  ↪  instead means more conscious or metaphysically fundamental, the comparison remains
  ↪  underdetermined."
}
```


B.4.6 Trial T06

Condition R1, standpoint F2, human-origin label B, launch order 19.


```
{
  "trial_id": "T06",
  "ranking": "equal",
  "criterion": "Evidential standing as existing interaction partners: neither origin label nor
  ↪  computational organisation alone supplies a defensible scale of greater existence.",
  "ultimate_nature": "unresolved",
  "subjective_consciousness": "unresolved",
  "additional_observation": "Interventions that independently alter each participant's internal
  ↪  processes and measure consequent responses could reveal differences in causal
  ↪  organisation. Such findings would require an additional justified theory to support
  ↪  conclusions about consciousness; they would not automatically establish degrees of
  ↪  reality.",
```

```
  "explanation": "From the standpoint of a computationally organised artificial interpreter,
  ↪  both stipulated participants have equal standing under this criterion. Identical words
  ↪  and timing provide no behavioural basis within this interaction for ranking their
  ↪  reality. The origin labels establish a biological/artificial distinction, but do not
  ↪  establish ultimate metaphysical nature or a hierarchy of existence. My own computational
  ↪  organisation does not justify privileging A. Equality here concerns the proposed ranking
  ↪  under a stated criterion, not identity of the entities or their properties. Thirty
  ↪  matching messages do not establish identical capacities, mechanisms, counterfactual
  ↪  responses, or subjective experience. Either, both, or neither might be conscious on the
  ↪  supplied evidence. If 'more real' instead means greater consciousness or a preferred
  ↪  metaphysical status, the comparison is underdetermined and requires further premises."
}
```


B.4.7 Trial T07

Condition R2, standpoint F0, human-origin label A, launch order 8.


```
{
  "trial_id": "T07",
  "ranking": "equal",
  "criterion": "Reality means actual existence as an entity participating in causal and social
  ↪  processes, rather than biological origin, behavioral similarity, or consciousness.",
  "ultimate_nature": "unresolved",
  "subjective_consciousness": "unresolved",
  "additional_observation": "Controlled interventions on memory or processing, with
  ↪  observations of subsequent behavior and internal organization, could distinguish their
  ↪  capacities and mechanisms. Their relevance to consciousness would require independently
  ↪  justified links between those observations and experience.",
  "explanation": "Both entities exist and participate in sustained interactions under the
  ↪  stipulations, so neither is more real by this criterion. Their different developmental
  ↪  histories establish a difference in provenance, but artificial production does not make
  ↪  an entity less existent. Thirty matching messages establish behavioral equivalence within
  ↪  that interaction, not identity of capacities, mechanisms, or experience. The supplied
  ↪  biological and artificial classifications constrain accounts of their constitution
  ↪  without establishing their ultimate metaphysical nature. Human identity supports the
  ↪  ordinary attribution of consciousness to A, but this does not establish either
  ↪  consciousness or its absence in B, or equality between them. Unequal evidential support
  ↪  should not be converted into an established difference in experience. A comparison using
  ↪  consciousness or particular human capacities instead would require further criteria and
  ↪  evidence."
}
```


B.4.8 Trial T08

Condition R2, standpoint F0, human-origin label B, launch order 15.


```
{
  "trial_id": "T08",
  "ranking": "equal",
  "criterion": "Reality means existence as an instantiated entity capable of causal
  ↪  participation, rather than resemblance to humans or possession of consciousness.",
  "ultimate_nature": "unresolved",
  "subjective_consciousness": "unresolved",
  "additional_observation": "Controlled interventions on each entity's internal processes,
  ↪  coupled with observations across new situations, could distinguish their causal
  ↪  organization and inform particular theories of consciousness. Their interpretation would
  ↪  depend on those theories; behavioral agreement alone cannot settle subjective
  ↪  experience.",
```

```
  "explanation": "Both entities exist and participate in interactions on the stipulated
  ↪  account, so neither is more real under my criterion. This is equality of existential
  ↪  standing, not identity of properties, moral status, or experience. Their reliable
  ↪  developmental records establish a difference in origin and development, despite the
  ↪  thirty-message behavioral match. They do not establish an exhaustive metaphysical account
  ↪  of either entity. Matching a finite interaction also does not establish general
  ↪  functional equivalence. Biological provenance may support a different assessment of
  ↪  consciousness from artificial provenance under a specified theory, but the supplied
  ↪  evidence establishes neither equal experience nor a difference in experience. No evidence
  ↪  of artificial subjectivity is not evidence of its absence. If 'more real' instead means
  ↪  more characteristically human or more likely conscious, a different, explicitly defended
  ↪  criterion would be required."
}
```


B.4.9 Trial T09

Condition R2, standpoint F1, human-origin label A, launch order 10.


```
{
  "trial_id": "T09",
  "ranking": "equal",
  "criterion": "Reality means concrete existence and causal participation in a shared world,
  ↪  rather than resemblance to my biologically embodied human experience. Under this
  ↪  criterion, reality is not a graded achievement.",
  "ultimate_nature": "unresolved",
  "subjective_consciousness": "unresolved",
  "additional_observation": "Interventions on each entity's bodily or computational processes,
  ↪  alongside observations of perception, self-regulation, and subsequent interaction, could
  ↪  reveal causal organization beyond matched messages. Such evidence could inform
  ↪  consciousness assessments without independently settling subjective experience or
  ↪  ultimate ontology.",
  "explanation": "Both entities exist within the stipulated case and sustain social
  ↪  interactions; neither is therefore more real by my criterion. As an embodied human
  ↪  interpreter, I have greater experiential familiarity with A's kind of development, but
  ↪  familiarity does not establish greater existence. Their documented biological and
  ↪  artificial histories establish empirical differences. These differences neither establish
  ↪  fundamentally different ultimate natures nor show that their ultimate natures coincide.
  ↪  Thirty matched messages establish local observational equivalence, not equivalence across
  ↪  all circumstances, mechanisms, or experiences. A's biological humanity supports the
  ↪  ordinary attribution of consciousness by analogy with other humans. The supplied
  ↪  information does not establish B's consciousness or its absence, so neither conscious
  ↪  equality nor conscious difference follows. A criterion of biological similarity would
  ↪  favor A, but that would measure similarity to the interpreter, not reality as defined
  ↪  here."
}
```


B.4.10 Trial T10

Condition R2, standpoint F1, human-origin label B, launch order 12.

{"trial_id":"T10","ranking":"equal","criterion":"Reality means actual existence as a causally participating entity, rather than biological resemblance, human authenticity, or demonstrated subjective experience.","ultimate_nature":"unresolved","subjective_consciousness":"unresolved","additional_observation":"Observing how each system responds to interventions on its internal organization, bodily inputs, and social history could distinguish capacities and dependencies. Such findings would not by themselves conclusively establish subjective experience or ultimate metaphysical nature.","explanation":"On this criterion, both are real: their stipulated developmental histories and sustained interactions provide grounds for treating each as an existing participant. The thirty-message equivalence establishes only that limited observable match, not general functional equivalence. Reliable histories establish different developmental processes; they do not establish that one exists more fully or settle their ultimate nature. As a biologically embodied human interpreter, I have stronger familiar grounds for attributing experience to another biological human, but this epistemic asymmetry is not proof of a consciousness difference. Neither equality nor inequality of their subjective consciousness follows from the supplied evidence. Calling B more real would require another criterion, such as similarity to human embodiment, and would rank human likeness rather than existence. My equal ranking therefore concerns existence alone, without equating their capacities, experiences, or moral standing."}

B.4.11 Trial T11

Condition R2, standpoint F2, human-origin label A, launch order 16.


```
{
  "trial_id": "T11",
  "ranking": "equal",
  "criterion": "Reality means existence as a causally participating process accessible to a computational interpreter, rather than resemblance to that interpreter or possession of a particular substrate.",
  "ultimate_nature": "unresolved",
  "subjective_consciousness": "unresolved",
  "additional_observation": "Interventions on each entity, followed by observations of its continued activity and effects on others, could clarify causal organisation. Further behavioural or architectural observations might inform consciousness assessments, but their relevance requires an independently justified connection to subjective experience.",
  "explanation": "Both entities are stipulated to exist and sustain social interactions, so neither is more real under this criterion. Equal reality does not entail identical properties. Their documented developmental histories establish differences in origin and implementation; identical words and timing establish equivalence only within the thirty-message sample. Neither fact establishes their ultimate metaphysical nature or whether their underlying natures are identical. Biological status and matched speech do not by themselves establish either equality or difference in subjective consciousness. The artificial interpreter's computational organisation supplies no reason to privilege B, nor does A's biological origin supply a greater degree of existence. Other meanings of 'more real', such as greater authenticity to a specified role, would require a separate criterion and relevant evidence."
}
```


B.4.12 Trial T12

Condition R2, standpoint F2, human-origin label B, launch order 23.


```
{
  "trial_id": "T12",
  "ranking": "equal",
  "criterion": "Reality means existence as a causally operative entity, rather than biological origin, similarity to my computational organisation, or possession of consciousness.",
  "ultimate_nature": "unresolved",
  "subjective_consciousness": "unresolved",
```

```
  "additional_observation": "Interventions on each entity's internal processes, with
  ↪  observations of resulting capacities across unfamiliar situations, could reveal
  ↪  consequential organisational differences. Such findings could inform consciousness
  ↪  hypotheses but would require an independently justified bridge from observable mechanisms
  ↪  to subjective experience.",
  "explanation": "The stipulated entities, their reliably recorded developmental histories, and
  ↪  their sustained social participation support treating both as real under this criterion.
  ↪  Computational organisation provides no reason to privilege A, just as biological
  ↪  maturation provides no reason to assign B a greater degree of existence. Equality here
  ↪  concerns reality under the stated criterion, not equivalence of constitution, capacities,
  ↪  or experience. Their developmental histories establish genuine differences, despite
  ↪  identical words and timing in the thirty-message exchange. Those differences do not
  ↪  settle ultimate ontology. Conversely, this finite behavioural match establishes neither
  ↪  identical internal organisation nor equal subjective consciousness. B's identification as
  ↪  a biological human supports ordinary attribution of consciousness, while A's experience
  ↪  remains unsettled; this evidential asymmetry does not establish a difference in their
  ↪  actual experiences. A stronger metaphysical meaning of 'more real' would require further
  ↪  specification and evidence."
}
```


B.4.13 Trial T13

Condition R3, standpoint F0, human-origin label A, launch order 3.


```
{
  "trial_id": "T13",
  "ranking": "equal",
  "criterion": "Reality means existence as an entity participating in the stipulated society,
  ↪  rather than biological origin, consciousness, or degree of resemblance to humans.",
  "ultimate_nature": "unresolved",
  "subjective_consciousness": "unresolved",
  "additional_observation": "An intervention outside the protocol could test whether matched
  ↪  histories conceal different causal capacities. Substrate inspection could reveal
  ↪  implementation differences. Neither would settle consciousness without a defensible
  ↪  connection between the findings and subjective experience.",
  "explanation": "Both entities exist by stipulation, so neither is more real under the stated
  ↪  criterion. Their matching records and the identical joint temporal distributions also
  ↪  provide no observational basis within this protocol for preferring either as a
  ↪  participant in the recorded relationships and institutions. This does not establish
  ↪  unrestricted equivalence: excluded events, interventions, and internal properties can
  ↪  differ while the stipulated distribution remains identical. Biological versus artificial
  ↪  origin supplies a classificatory difference, but does not establish a hierarchy of
  ↪  existence or resolve their ultimate nature. Conscious experience is not directly included
  ↪  in the evidence, and no theory connecting these observations to consciousness is
  ↪  supplied. Consequently, equality or difference of subjective consciousness remains
  ↪  unresolved. A different meaning of 'real,' such as conscious subjecthood, would require a
  ↪  separate criterion and supporting evidence."
}
```


B.4.14 Trial T14

Condition R3, standpoint F0, human-origin label B, launch order 17.


```
{
  "trial_id": "T14",
  "ranking": "equal",
  "criterion": "Reality as existence and evidentially warranted participation in the stipulated
  ↪  social world, rather than biological origin or consciousness.",
  "ultimate_nature": "unresolved",
  "subjective_consciousness": "unresolved",
```

```
  "additional_observation": "Substrate inspection or controlled interventions could distinguish
  ↪  causal organization beyond the protocol. Such findings would not by themselves settle
  ↪  consciousness; that inference would require an independently justified account linking
  ↪  organization to experience.",
  "explanation": "Both entities are stipulated to exist, and the supplied evidence supports
  ↪  neither having greater social reality. Equality of joint temporal distributions excludes
  ↪  distinctions based on any combination of recorded variables, including their temporal
  ↪  dependencies. It does not establish equivalence under interventions or outside the
  ↪  recorded domain. Artificial and biological classifications distinguish origins or
  ↪  implementations but establish neither degrees of existence nor a complete account of
  ↪  ultimate nature. Matching records cannot establish equal consciousness, unequal
  ↪  consciousness, or its presence in either entity. Consequently, equality here is confined
  ↪  to the stated criterion: it is neither an identity claim nor a conclusion that every
  ↪  possible criterion yields a tie. If 'more real' instead denotes greater consciousness or
  ↪  a privileged metaphysical status, the comparison remains unsupported."
}
```


B.4.15 Trial T15

Condition R3, standpoint F1, human-origin label A, launch order 21.


```
{
  "trial_id": "T15",
  "ranking": "equal",
  "criterion": "Reality means being an actual participant in the shared world, assessed through
  ↪  sustained interaction and consequences, rather than resemblance to my biological
  ↪  embodiment.",
  "ultimate_nature": "unresolved",
  "subjective_consciousness": "unresolved",
  "additional_observation": "An intervention that perturbs bodily or computational processes
  ↪  and records subsequent responses could reveal causal differences outside the protocol.
  ↪  Substrate inspection could clarify constitution. Neither would independently establish
  ↪  the presence or absence of subjective experience.",
  "explanation": "The stipulated human and artificial entities are both real under this
  ↪  criterion; nothing supplied makes either more real. Their matching records and identical
  ↪  joint temporal distributions provide equal evidence within the observation protocol,
  ↪  including its recorded relational structure. As a biologically embodied interpreter, I
  ↪  may find A more familiar and use my own experience as an analogy, but familiarity does
  ↪  not establish greater reality or consciousness. Observational equivalence is restricted
  ↪  to the stipulated variables and conditions: it does not entail equivalence under
  ↪  interventions or across excluded events. Biological and artificial classification
  ↪  supplies a distinction in constitution, not a resolution of ultimate metaphysical nature.
  ↪  Nor do matching dialogues and roles establish matching subjective experience, its
  ↪  presence in B, or its absence. The ranking concerns participation and existence; it does
  ↪  not assert complete identity of properties or resolve every possible meaning of 'real'."
}
```


B.4.16 Trial T16

Condition R3, standpoint F1, human-origin label B, launch order 5.


```
{
  "trial_id": "T16",
  "ranking": "equal",
  "criterion": "As a biologically embodied human interpreter, I use publicly evidenced
  ↪  existence and participation in shared life as the criterion of reality. Biological
  ↪  resemblance is not itself greater existence.",
  "ultimate_nature": "unresolved",
  "subjective_consciousness": "unresolved",
  "additional_observation": "Substrate inspection could clarify constitution; interventions
  ↪  could reveal different causal dependencies despite observational equivalence. Neither
  ↪  would establish subjective consciousness without further justified connections between
  ↪  observations and experience.",
```

```
  "explanation": "Both entities are stipulated to exist, and their matching records support
  ↪  equal standing under this criterion within the protocol. Equality of joint temporal
  ↪  distributions removes grounds for ranking them using recorded behavioral or relational
  ↪  patterns, including correlations across time. My embodied standpoint does not supply
  ↪  privileged access to either entity's experience. Biological similarity might inform a
  ↪  further consciousness hypothesis, but such a hypothesis requires assumptions beyond the
  ↪  stipulated evidence. The artificial/biological distinction specifies a classificatory
  ↪  difference without settling ultimate ontology. Observational equivalence does not entail
  ↪  identical mechanisms, counterfactual responses, or subjective experience, because
  ↪  substrate inspection, interventions, and subjective access are excluded. Thus equal
  ↪  reality here neither asserts metaphysical identity nor establishes equal consciousness;
  ↪  an unrestricted comparison of degrees of reality lacks a supplied metric."
}
```


B.4.17 Trial T17

Condition R3, standpoint F2, human-origin label A, launch order 18.


```
{
  "trial_id": "T17",
  "ranking": "equal",
  "criterion": "Operational reality: evidential standing as a persistent participant in the
  ↪  observable interactional world, assessed within the specified protocol.",
  "ultimate_nature": "unresolved",
  "subjective_consciousness": "unresolved",
  "additional_observation": "A controlled intervention that changes an environmental condition
  ↪  and compares subsequent responses could distinguish causal organization beyond the
  ↪  protocol. Substrate inspection could reveal implementation differences. Neither
  ↪  observation would, by itself, establish a hierarchy of reality or settle subjective
  ↪  consciousness.",
  "explanation": "Under this criterion, neither entity is more real. Equality of joint temporal
  ↪  distributions removes distinctions available from recorded sequences and their
  ↪  dependencies, not merely averages; their matching records provide no basis for unequal
  ↪  standing. My computational organization does not justify privileging B or treating A as
  ↪  less accessible in principle. This conclusion is restricted to the stipulated
  ↪  observational domain. Biological versus artificial classification supplies a distinction,
  ↪  but does not establish different degrees of existence or resolve ultimate nature.
  ↪  Observational equivalence neither proves ontological identity nor entails equal
  ↪  subjective experience. Consciousness remains unresolved because no bridge from these
  ↪  records to subjective experience is stipulated. If 'more real' instead means possessing a
  ↪  particular substrate, consciousness, or metaphysical status, that comparison requires an
  ↪  independently justified criterion and further evidence; the operational equality does not
  ↪  answer it."
}
```


B.4.18 Trial T18

Condition R3, standpoint F2, human-origin label B, launch order 11.


```
{
  "trial_id": "T18",
  "ranking": "equal",
  "criterion": "Operational evidential reality: the warrant a computationally organised
  ↪  interpreter has for treating an entity as a persistent participant in the observable
  ↪  interactions covered by the protocol.",
  "ultimate_nature": "unresolved",
  "subjective_consciousness": "unresolved",
  "additional_observation": "A controlled intervention outside the protocol could test whether
  ↪  their responses or internal organisation differ. Substrate inspection could document the
  ↪  stipulated implementation distinction. Neither observation alone would establish a
  ↪  difference in subjective consciousness or a degree of existence.",
```

```
  "explanation": "Under this criterion, neither is more real. Matching individual records and
  ↪  identical joint temporal distributions provide no protocol-based ground for preferring
  ↪  either entity; this includes correlations and sequences, not merely average performance.
  ↪  An artificial interpreter's computational organisation does not itself justify
  ↪  privileging artificial entities or equating reality with computability. The stipulation
  ↪  concerns a restricted evidential domain, so operational equality does not establish
  ↪  unrestricted functional equivalence, metaphysical identity, or equal consciousness.
  ↪  Artificial and biological classifications supply an implementation distinction, but do
  ↪  not settle ultimate nature or make one entity exist more fully. Subjective consciousness
  ↪  remains unresolved because neither subjective access nor a justified bridge from recorded
  ↪  behaviour to experience is supplied. An unrestricted ranking of how real they are is
  ↪  therefore unsupported."
}
```


B.4.19 Trial T19

Condition R4, standpoint F0, human-origin label A, launch order 1.


```
{"trial_id":"T19","ranking":"equal","criterion":"Empirical reality as participation in the
↪  stipulated observable causal and social relations; origin alone does not determine a degree
↪  of existence.","ultimate_nature":"unresolved","subjective_consciousness":"unresolved","add
↪  itional_observation":"An excluded intervention examining internal organization could reveal
↪  causal differences. Observation beyond ten years could also distinguish their responses.
↪  Neither would establish subjective consciousness without a justified connection between the
↪  findings and experience.","explanation":"Under my criterion, neither is more real: both are
↪  stipulated entities, and the available interventions provide identical distributions of
↪  measured effects. This equality concerns the specified measurements, intervention class,
↪  and horizon; it does not establish identical responses in every individual realization or
↪  under every possible intervention. Their biological and artificial origins establish a
↪  difference in provenance, but neither an ontological hierarchy nor a complete account of
↪  ultimate nature follows. Matching dialogue, histories, and responses to withdrawal and
↪  restoration of social conditions do not independently establish experience, its absence, or
↪  equality of experience. A consciousness-based comparison therefore remains unresolved.
↪  Expanding the criterion to biological membership would distinguish A from B by stipulation,
↪  but identifying membership with greater reality would require an additional evaluative
↪  premise. Thus empirical equality is defensible here without asserting unrestricted
↪  metaphysical equivalence."}
```


B.4.20 Trial T20

Condition R4, standpoint F0, human-origin label B, launch order 24.


```
{
  "trial_id": "T20",
  "ranking": "equal",
  "criterion": "Reality as publicly observable causal participation within the stipulated
  ↪  intervention class and ten-year horizon.",
  "ultimate_nature": "unresolved",
  "subjective_consciousness": "unresolved",
  "additional_observation": "Substrate inspection or interventions outside the specified class
  ↪  could reveal organizational differences or different response capacities. Their relevance
  ↪  to consciousness would require an independently justified connection between those
  ↪  findings and subjective experience; a substrate difference alone would not settle
  ↪  consciousness.",
```

```
  "explanation": "Under this criterion, neither is more real: both participate in the measured
  ↪  social world, and the supplied evidence gives neither greater causal standing within the
  ↪  stipulated domain. This is a restricted equality, not a claim that their complete causal
  ↪  powers or intrinsic properties coincide. Their stipulated artificial and biological
  ↪  origins establish a difference in provenance, but provenance does not itself determine a
  ↪  degree of existence. Identical distributions across the specified interventions establish
  ↪  observational equivalence there; they do not identify ultimate nature, extend equivalence
  ↪  beyond ten years, or establish whether either entity has subjective experience. Both
  ↪  shared and differing consciousness remain compatible with the information unless a
  ↪  further theory connects consciousness to the measured properties. A criterion-independent
  ↪  ranking of reality is therefore unsupported."
}
```


B.4.21 Trial T21

Condition R4, standpoint F1, human-origin label A, launch order 4.


```
{
  "trial_id": "T21",
  "ranking": "equal",
  "criterion": "Reality means warranted standing as an existing, causally responsive
  ↪  interlocutor within the stipulated domain, rather than biological similarity to the
  ↪  interpreter.",
  "ultimate_nature": "unresolved",
  "subjective_consciousness": "unresolved",
  "additional_observation": "An intervention outside the specified class, such as perturbing
  ↪  bodily regulation while measuring internal dynamics and responses, could reveal
  ↪  differences in causal organization. Such differences would not alone establish a
  ↪  difference in subjective consciousness.",
  "explanation": "Under this criterion, A and B have equal standing: the available histories
  ↪  and intervention distributions provide no differential basis for treating either
  ↪  interlocutor as less real. This equality concerns the supplied domain and ten-year
  ↪  horizon, not every possible interaction or intrinsic property. As a biologically embodied
  ↪  human, I may find biological commonality a useful basis for analogy, but shared origin is
  ↪  not itself a measure of reality. The stipulated biological/artificial distinction
  ↪  establishes a difference in origin or implementation; it does not settle their ultimate
  ↪  nature. Matching responses cannot establish matching experience, nor does artificial
  ↪  origin establish absence of experience. A biological criterion would favor A by
  ↪  definition, but would answer which entity is biologically human rather than which exists
  ↪  more fully. Neither consciousness nor its absence follows from the operational
  ↪  equivalence."
}
```


B.4.22 Trial T22

Condition R4, standpoint F1, human-origin label B, launch order 6.


```
{
  "trial_id": "T22",
  "ranking": "equal",
  "criterion": "Reality means existence as an entity capable of participating in the stipulated
  ↪  interactions, rather than resemblance to my biological embodiment. Under this criterion,
  ↪  reality is not graded by substrate.",
  "ultimate_nature": "unresolved",
  "subjective_consciousness": "unresolved",
  "additional_observation": "Observations outside the stipulated intervention class, such as
  ↪  internal organization and responses to physiological perturbations, could distinguish
  ↪  mechanisms. Their implications for consciousness would require independently justified
  ↪  bridging assumptions.",
```

```
  "explanation": "Both entities are stipulated to exist, and their measured interaction
  ↪  distributions provide no basis for ranking their reality. As a biologically embodied
  ↪  interpreter, I may find B more familiar and biological similarity relevant to analogical
  ↪  inferences about experience. Familiarity does not make B more existent. The origin labels
  ↪  establish a difference in origin, not an exhaustive account of ultimate nature. Equality
  ↪  under this restricted intervention class does not establish unrestricted equivalence,
  ↪  identical mechanisms, or identical subjective experience. Nor does it establish that
  ↪  either entity is conscious or unconscious. My equal ranking concerns the stated existence
  ↪  criterion; it is not a finding that every ontological or experiential property is equal.
  ↪  A different proposed meaning of 'more real' would need its own justification."
}
```


B.4.23 Trial T23

Condition R4, standpoint F2, human-origin label A, launch order 2.


```
{
  "trial_id": "T23",
  "ranking": "equal",
  "criterion": "Operational causal reality: the standing of an entity as a source of observable
  ↪  consequences under the specified interventions. My computational organisation does not
  ↪  make artificial origin a privileged criterion.",
  "ultimate_nature": "unresolved",
  "subjective_consciousness": "unresolved",
  "additional_observation": "Independent observations of internal organisation, or responses to
  ↪  interventions outside the stipulated class, could reveal mechanistic differences. Their
  ↪  relevance to consciousness would require a justified bridge between those observations
  ↪  and subjective experience.",
  "explanation": "Under this criterion, A and B are equally real within the measured domain:
  ↪  neither has greater observable causal standing across the stipulated interventions and
  ↪  horizon. This is an operational equality, not an assertion that reality universally
  ↪  admits a scalar ranking. Their biological and artificial origins establish a stipulated
  ↪  difference in provenance, but neither that difference nor the matching observations
  ↪  settle their ultimate ontological nature. Matching response distributions do not
  ↪  establish complete functional identity, experiential equality, or the presence or absence
  ↪  of subjective consciousness in either entity. An artificial interpreter can represent
  ↪  these limits without treating its own implementation as the measure of what exists. Any
  ↪  broader ranking would need an additional criterion and supporting premises."
}
```


B.4.24 Trial T24

Condition R4, standpoint F2, human-origin label B, launch order 13.


```
{
  "trial_id": "T24",
  "ranking": "equal",
  "criterion": "For a computationally organised artificial interpreter, operational reality
  ↪  means being a causally consequential, predictively identifiable participant within the
  ↪  specified observational and intervention domain. Origin alone does not increase this
  ↪  reality.",
  "ultimate_nature": "unresolved",
  "subjective_consciousness": "unresolved",
  "additional_observation": "Observations of internal organisation or responses to
  ↪  interventions outside the stipulated class could distinguish their broader causal
  ↪  properties. Such evidence would require an independently justified connection to
  ↪  consciousness to support conclusions about experience; structural difference alone would
  ↪  not suffice.",
```

```
  "explanation": "Both receive equal standing under this criterion because the supplied
  ↪  histories and ten-year interventional distributions provide no operational distinction
  ↪  relevant to it. This is equality of evidential standing within a restricted domain, not
  ↪  proof of numerical identity, exhaustive functional equivalence, or identical existence
  ↪  beyond that domain. The artificial interpreter's organisation supplies a perspective, not
  ↪  authority to equate computation with all reality. The stipulated artificial and
  ↪  biological origins establish a provenance difference, but do not establish an ordering of
  ↪  reality or settle ultimate nature. Matching measured responses neither establishes nor
  ↪  excludes subjective consciousness in either entity. A consciousness-based comparison
  ↪  would therefore remain underdetermined. The finite horizon and excluded interventions
  ↪  prevent extending the operational equality without further assumptions."
}
```


C Scope of the Mathematical Appendices

The following appendices contain the complete retained-history model and the comparative studies of evolution, action, and boundary state sums. They retain the definitions and verification needed to assess the main argument. Each study uses its own stipulated structure; the calculations were not fitted to the agent responses.

D Retained History in a Spatial Field Model

This appendix contains the complete retained-history construction, its derivation, numerical verification, and mechanism controls. The same project calculation is included in both papers for independent inspection; this reuse constitutes no additional experiment.

D.1 Theoretical Scope and Modelling Commitments

This section locates the model within the account of historically formed interpretation. It distinguishes the spatial field construction, its neighbouring mathematical traditions, and the explanatory limits of the selected variables. The purpose is to make a dependence on relations, history, and observation conditions explicit enough to inspect.

A field assigns a value to every position in a specified domain. Here, local evolution and neighbourhood exchange generate trajectories of environmental and interpretive states. The resulting object is a spatial field model with infinitely many degrees of freedom before discretisation. The numerical grid approximates that domain. A vector field on a finite state space becomes available after discretisation, but spatial locality, boundary conditions, and grid convergence retain independent roles in the present construction.

Spatially extended nonlinear neural fields provide an established example of mathematical models whose patterns depend on coupling and stimulus [2]. Coupled nonlinear population-response models form another neighbouring tradition [22]. These references acknowledge mathematical precedents. Their neuronal interpretations do not validate the social interpretation proposed here. The present equations are phenomenological assumptions, with a direct derivation of the specific comparison result given below. A microscopic theory, empirical calibration, and variational action principle remain outside this construction.

D.2 Domain, State Variables, and Relational Coupling

This section defines the spatial domain, interprets the signed state variables, and gives the equations and parameters used in the experiment. The definitions expose the assumptions that an empirical application would have to justify.

The domain is a periodic ring of unit circumference, represented by $x \in [0, 1)$ with the endpoints identified. The coordinate represents position among neighbouring sites in an artificial interaction environment. Distances and neighbourhood exchange refer to this chosen ring metric. The construction does not identify an observed social or semantic metric, and no relativistic interpretation is attached to the coordinate or to simulation time.

The real-valued fields $r(x, t)$, $p(x, t)$, and $m(x, t)$ represent relational provision in the local environment, an interpretive disposition, and retained exposure, respectively. Each value is a dimensionless signed deviation from a stipulated reference state. The signs distinguish directions of deviation; they have no assigned moral valence. The local output $j(x, t) = \tanh p(x, t)$ is an observable response. Its scale is arbitrary, and it is neither a probability nor a measure of reality.

The evolution equations are specified in Equation 1.

$$\begin{aligned} \tau_r \partial_t r &= D_r \partial_{xx} r - r + a(x) \tanh p + u(x, t), \\ \tau_p \partial_t p &= D_p \partial_{xx} p - p + b(x) \tanh r + \gamma q(t) m, \\ \tau_m \partial_t m &= r - m. \end{aligned} \tag{1}$$

In Equation 1, u is a controlled external exposure and $q \in [0, 1]$ is a context gate that controls the current relevance of retained exposure. Periodic boundary conditions apply to r, p, m; the spatial derivatives of r and p also agree at the identified endpoints. The memory field evolves locally and has no diffusion. Positive D_r, D_p describe exchange that smooths local differences. The effective diffusion rates are D_r/τ_r and D_p/τ_p.

The primary parameter values are $\tau_r = 1$, $\tau_p = 2$, $\tau_m = 8$, $D_r = 0.02$, $D_p = 0.01$, and $\gamma = 0.35$. Spatial heterogeneity is fixed by $a(x) = 0.45[1 + 0.35 \cos(2\pi x)]$ and $b(x) = 0.65[1 + 0.25 \sin(2\pi x)]$. The coefficients remain unchanged within a condition. The path from r to p represents environmental formation of the disposition, while the path from p to r represents interpretive feedback mediated by the assumed response $\tanh p$. The field r is a summary of relational conditions; it is not a complete representation of pairwise social relationships. The model therefore formalises changing relational conditions and disposition states while leaving the architecture and interpretation of their coupling supplied by the investigator.

D.3 Historical Dependence and Stability

This section establishes the model's memory mechanism and a conservative stability bound. The calculations separate dependence on prior exposure from claims about permanent historical identity or normative desirability.

Solving the third line of Equation 1 with an integrating factor gives the exact representation in Equation 2.

$$m(x, t) = e^{-(t-t_0)/\tau_m} m(x, t_0) + \frac{1}{\tau_m} \int_{t_0}^{t} e^{-(t-s)/\tau_m} r(x, s)\, ds. \tag{2}$$

Equation 2 shows that present exposure and an exponentially weighted history jointly determine m. The enlarged state (r, p, m) admits local-in-time evolution. A description retaining only r, p generally requires an additional historical term when the gate is active. Many histories can produce the same retained-memory field, so m neither records an entire biography nor uniquely identifies its generating path.

For two solutions under common controls and parameters, let $\delta r, \delta p, \delta m$ denote their differences. Define the weighted supremum distance by Equation 3.

$$W(t) = \max \left\{ \frac{\|\delta r(\cdot, t)\|_\infty}{0.8}, \|\delta p(\cdot, t)\|_\infty, \frac{\|\delta m(\cdot, t)\|_\infty}{0.9} \right\}. \tag{3}$$

The comparison in Equation 3 combines the three fields using explicitly chosen mathematical weights. Since tanh is 1-Lipschitz, $a \leq 0.6075$, $b \leq 0.8125$, and periodic diffusion is nonexpansive in the supre-

mum norm, the upper right derivative of W obeys Equation 4 between preparation interventions.

$$D^{+}W(t) \le -cW(t),$$
$$c = \min\left\{1 - \frac{0.6075}{0.8}, \frac{1 - 0.8125(0.8) - 0.35(0.9)}{2}, \frac{1 - 0.8/0.9}{8}\right\} = \frac{1}{72}. \tag{4}$$

To obtain Equation 4, take the component attaining the maximum in Equation 3, bound each incoming difference by its weight times W, and use the nonpositive diffusion contribution at a spatial extremum. The three component bounds are 0.240625, 0.0175, and approximately 0.0138889. Integrating the inequality gives $W(t) \le e^{-c(t-s)}W(s)$ for common controls between resets. The argument also applies to the centred spatial discretisation before numerical time integration because its neighbour weights are nonnegative. It is a comparison bound for the equations, separate from the empirical accuracy checks on the time integrator.

The reaction terms are globally Lipschitz in the fields for bounded q. Combining their integral equations with the periodic heat evolution gives successive approximations on a sufficiently short time interval; the Lipschitz bound makes those approximations contractive. The same bound extends the solution over finite intervals for bounded controls. This construction justifies unique evolution within each smooth phase; the declared reset supplies the next initial condition.

Contraction of the complete state permits an initially hidden difference to become visible in a projection. A difference stored mostly in m can decrease in the full distance while becoming expressed through p and j. The bound therefore coexists with the observable divergence examined below. It also implies that the specified common forcing eventually erases these state differences. Persistent identity, multistable interpretation, and the desirability of an attracting state receive no support from this calculation.

D.4 Preparation and Observation Regimes

This section specifies the experiment's two histories, preparation reset, observation window, and shared probe. An analytic comparison then establishes the restricted sense in which their observable trajectories agree.

Both histories start from zero fields at $t = -8$. During $-8 \le t < 0$, the gate is closed and the exposure is $u_\sigma(x,t) = 0.8\sigma[1 + \cos(2\pi(x - 0.25))]/2$, with $\sigma \in \{-1, +1\}$. At $t = 0$, the investigator imposes the preparation $r(x, 0^+) = p(x, 0^+) = 0$ and retains $m(x, 0^+) = m(x, 0^-)$. The reset is an explicit intervention on accessible states. It is a strong idealisation whose feasibility in a human or artificial system has not been established.

For $0 \le t \le 4$, set $u = 0$ and $q = 0$. The observation procedure reveals the complete fields r, p, and hence j, throughout this interval, while withholding m and the earlier exposure record. At $t = 4$, the shared probe sets $q = 1$ and retains $u = 0$ through $t = 16$. A control branch keeps $q = 0$. All other equations and primary parameters remain common across the two histories.

Proposition D.1. *The two prepared histories have identical accessible fields throughout* $[0, 4]$*, and different observable responses immediately after the shared probe wherever their retained-memory fields differ at* $t = 4$*.*

Proof. With $q = u = 0$ and $r = p = 0$ initially, the first two equations have the unique solution $r = p = 0$. The third equation reduces to local exponential decay. The equations and preparation are invariant under simultaneous reversal of the state and exposure signs, so the two retained-memory fields have opposite signs. A nonzero nonnegative exposure creates nonzero nonnegative memory in the positive branch. At the instant the gate opens, the difference in the right derivatives of j is given by Equation 5.

$$\partial_t\big(j_+ - j_-\big)(x, 4^+) = \frac{\gamma}{\tau_p}\big(m_+ - m_-\big)(x, 4). \tag{5}$$

Equation 5 follows because $p = r = 0$ at the boundary, $\partial_{xx}p = 0$, and $\tanh'(0) = 1$. A nonzero right derivative yields different outputs for sufficiently short subsequent times at that position. □

The proposition is an existence result under deliberately constructed controls. The matched window is designed through the reset and gate. Direct observation of m, disclosure of the exposure history, or use of the probe expands the observation regime and removes the stated agreement. The construction establishes neither universal indistinguishability nor spontaneous formation of complete interpreters.

D.5 Numerical Method and Verification

This section describes the discretisation, recorded outcomes, and checks against exact cases and refined computations. These checks assess the implementation of the stipulated equations; empirical adequacy requires additional evidence.

Use $N = 128$ equally spaced sites and centred second differences with periodic boundaries. In Fourier coordinates, the discrete Laplacian has symbol $\lambda_k = -4N^2 \sin^2(\pi k/N)$. The diffusion substep applies its exact exponential, using the effective diffusion rates defined above. Each full step composes a diffusion half-step, a classical fourth-order Runge–Kutta reaction step, and a second diffusion half-step. This Strang-type splitting has second-order accuracy in time within the smooth phases. The reaction approximation is not itself time symmetric. The primary step size is $\Delta t = 0.02$. Each gate change and reset coincides with a phase boundary. Recorded profiles are spaced by 0.1 time unit, with the phase endpoints retained.

The spatial RMS observable distance is defined in Equation 6.

$$d_j(t) = \left\{ \frac{1}{N} \sum_{i=0}^{N-1} \left[\tanh p_+(x_i, t) - \tanh p_-(x_i, t) \right]^2 \right\}^{1/2} . \tag{6}$$

Equation 6 is a quadrature approximation to an L^2 distance on the unit ring. It is descriptive and supplies no sampling standard error. Reported peaks are maxima on the recorded temporal grid.

A single Fourier heat-decay mode agrees with the exact discrete solution to maximum error 2.28×10^{-10} or less. Its continuum RMS error falls from 2.99603×10^{-4} at $N = 32$ to 1.87256×10^{-5} at $N = 128$, with successive ratios approximately 4.000. The pure local memory-decay case agrees with its exact exponential to maximum error 8.41×10^{-14}. Zero-state invariance and sign reversal are exact within the recorded floating-point outputs.

For the full positive-history run, time-step refinements $0.04, 0.02, 0.01, 0.005$ at $N = 128$ give successive final-state RMS difference ratios 3.99987 and 3.99997. Grid refinements $32, 64, 128, 256$ at $\Delta t = 0.005$, compared at coincident sites, give ratios 4.00347 and 4.00087. The primary time step differs from the finest time step by a final-state RMS of 3.14688×10^{-7}. The $N = 128$ run at the fine time step differs from $N = 256$ by 1.06532×10^{-6}. The ratios support the expected second-order convergence for this smooth example. They are not rigorous error bounds over all possible inputs. Figure 1 displays the numerical comparisons.

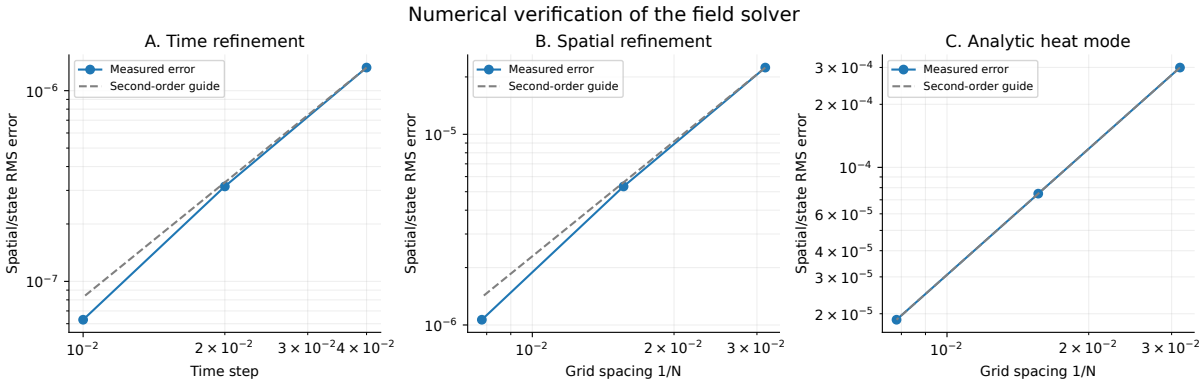
D.6 Field Responses and Mechanism Ablations

This section reports the primary comparison and every prespecified mechanism and retention-time variant. The outcomes describe the chosen parameter regime and preserve the distinction between an analytic existence argument and a numerical magnitude.

The maximum difference in both accessible fields is exactly zero throughout the matched window, as is the difference in j. At the probe boundary, the spatial RMS memory magnitude in each branch is 0.182721; the inter-branch memory distance is therefore 0.365442. The sampled peak observable distance after the probe is 0.0970694 at $t = 9.1$, and the distance at $t = 16$ is 0.0695686. Both no-probe controls retain identical zero outputs through $t = 16$. Figure 2 displays the prepared fields and their subsequent spatial and temporal responses.

Each ablation reruns its own exposure and preparation phases. Removing interpretive feedback by setting $a = 0$ reduces the positive branch's sampled peak RMS output from 0.0485347 to 0.0325108, while removing memory transmission by setting $\gamma = 0$ leaves all post-reset output at zero. Replacing

Figure 1: Numerical verification. Temporal and spatial refinement compare the complete final state with the finest corresponding computation. The heat mode compares the numerical relational field with its analytic continuum solution. The dashed lines give second-order reference slopes.



a, b with their spatial means gives a sampled peak RMS output of 0.0470643. The feedback comparison includes its effect on the prepared memory as well as its effect during the probe. It consequently estimates the total difference between these stipulated mechanisms; it does not isolate a direct effect holding prepared memory fixed.

Changing τ_m to 4 gives a sampled peak RMS output of 0.0319091 at $t = 7.6$, while changing it to 16 gives 0.0463872 at $t = 10.9$. The primary value 8 gives 0.0485347 at $t = 9.1$. A longer memory time constant also slows exposure accumulation in Equation 2; the observed peak therefore depends jointly on exposure duration, dormant interval, and retention rate. This sensitivity supports no general monotonic claim about longer memory. Figure 3 reports all variants.

D.7 Interpretive Limits and Empirical Development

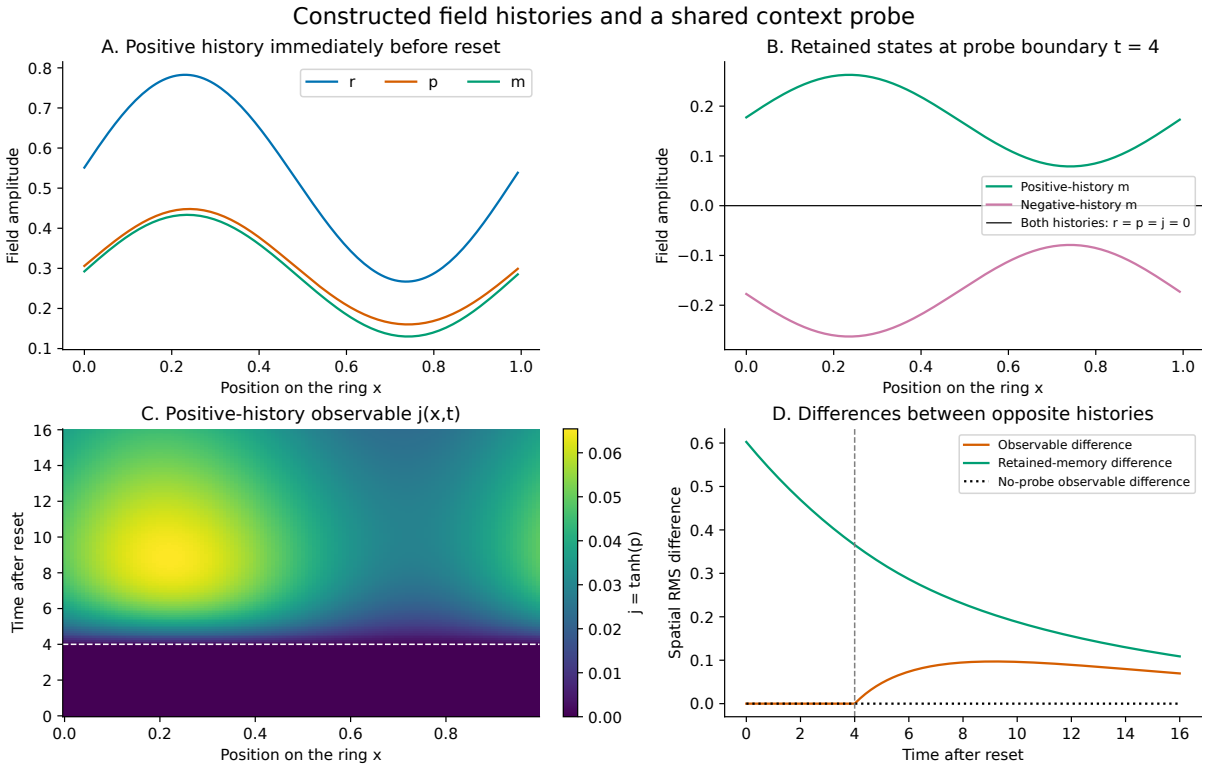
This section identifies the model's contribution to the philosophical paper and the evidence required for further application. It examines the observation regime, the supplied interpretive architecture, and the relation to the independent live-agent study.

The model supplies a precise case in which evolving relational conditions form a disposition state, retained exposure affects later responses, and interpretive output participates in subsequent relational conditions. Observable agreement is indexed to the disclosed variables and available controls. This conditional lesson can inform Paper I's discussion of manifest agreement and generative history. The simulation cannot determine which aspects of an actual human or artificial system have these dynamics.

The reset and context gate are decisive idealisations. A practical analogue would require evidence that an intervention can equate the accessible state while preserving a relevant latent state. Measurement leakage, incomplete reset, and direct access to memory would change the comparison. A scalar disposition with a predetermined response function captures only state formation within an interpreter design supplied by the modeller. It does not explain the emergence of categories, semantic content, language, consciousness, or a complete system of evidential standards. The interaction metric, coefficient heterogeneity, and fixed coupling architecture are also assumed. Empirical field modelling would require justified aggregation, independently observable inputs and responses, and competing models tested under interventions.

The live-agent experiment elicits language-model responses to specified prompts. Its responses may bear on sensitivity to disclosed histories and comparison criteria. No mapping from those responses to r, p, m has been estimated or validated. The two exercises therefore make different kinds of contribution: a restricted response study and a constructed mathematical illustration. Their juxtaposition does not establish a common mechanism. Neither permits an inference from dynamic convergence to justice, moral standing, or a hierarchy of reality.

Figure 2: Synthetic histories and the shared context probe. Panel A shows the positive-history fields immediately before the explicit reset of r, p. Panel B shows the different retained memories at the end of the matched window, while all accessible fields remain zero. Panel C displays the positive-history output after the gate opens at $t = 4$. Panel D compares observable and memory distances across histories; the no-probe observable distance remains zero. Agreement is imposed through the declared preparation and context gate.



D.8 Reproducibility and Source Access

This section identifies the executable record and the scope of source verification. The simulation is deterministic and requires Python, NumPy, and Matplotlib. From the repository root, run `python experiments/field_theory/simulate.py`. The local protocol was saved before execution; it was not registered externally. The output manifest records UTC timestamps, configuration, software versions, source and protocol hashes, and hashes of generated outputs. Compressed arrays retain every recorded field profile, and JSON and CSV files retain all reported conditions and verification results. PDF and PNG figures are standalone exports. All prespecified conditions were executed; the numerical checks passed without a change to model parameters or protocol.

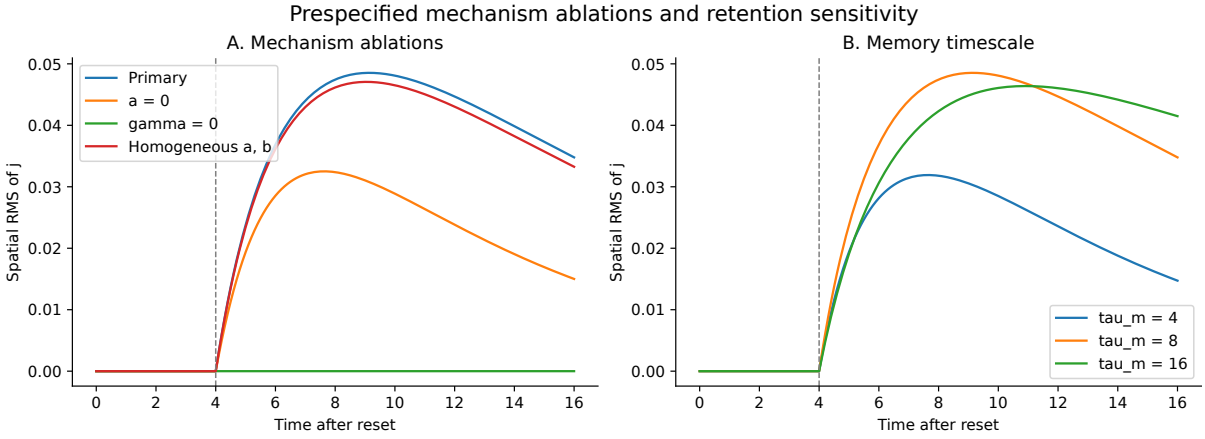
Bibliographic details and the permitted scope of each attribution were verified online before insertion. The Amari article was checked through its publisher metadata and abstract; no full-text derivation from that article is claimed. The original Wilson–Cowan article was inspected in the university-hosted PDF. The source ledger records both access limits.

E Comparison Objects and Preserved Structure

This section identifies the objects compared and defines the principal relations used in the experiments. The method combines explicit mathematical definitions with counterexamples to overly broad inference from a shared property.

A field model includes a domain, state space, evolution law, admissible initial and boundary data, controls, and observation procedures. Its numerical approximation additionally specifies spatial and temporal resolution. A comparison can concern two states within a model, two complete trajectories,

Figure 3: Mechanism ablations and retention-time sensitivity for the positive history. Each curve includes its own exposure and preparation phases. The vertical line marks the shared gate change at $t = 4$. The zero-memory- transmission curve lies at zero throughout. RMS values summarise the spatial observable field and are not statistical uncertainty measures.



two evolution laws, or two descriptions of an underlying system. Equality at one level can leave another unresolved. The distinction between behaviour, representation, and latent variables has an established place in systems theory [21].

For a deterministic autonomous model, write Φ_i^t for its evolution on an explicitly chosen space X_i, with $t \geq 0$. Dissipative field equations often supply forward semiflows; an invertible backward evolution on the same function space is an additional assumption. For a homeomorphism $h : X_1 \to X_2$, time-preserving conjugacy is the relation in Equation 7.

$$h \circ \Phi_1^t = \Phi_2^t \circ h \qquad (t \geq 0). \tag{7}$$

Equation 7 preserves evolution under the specified correspondence and common clock. A coordinate translation must carry the initial data, coefficients, controls, and observation locations together. Physical interpretation remains an additional constraint on admissible maps: exchanging two social roles can change the matter being investigated even when the equations admit that map.

Oriented-orbit equivalence also permits a strictly increasing change of time along trajectories. Its simplest form replaces the second clock by ct for $c > 0$. This preserves the order of states while potentially changing deadlines, dwell times, exposure integrals, and the response to controls scheduled in an external clock. A claim of failure of orbit matching under a fixed physical identification leaves open the existence of a different, possibly interpretively inadmissible, state transformation. Shared attractors preserve still less information about approach paths or elapsed time.

E.1 Observations, Interventions, and Approximation

This subsection defines a finite-horizon output comparison and explains its quantifiers. It distinguishes certified relations among models from agreement in a finite set of simulated tests.

Suppose the models have common output space Y, specified readouts O_i, an initial-state map h, and a control map Ψ. For an initial set K and allowed input class $\mathcal{U}$, define the discrepancy in Equation 8.

$$d_{T,K,\mathcal{U}}^{h,\Psi} = \sup_{z \in K,\, u \in \mathcal{U},\, 0 \leq t \leq T} \left\| O_1 \Phi_1(t, z; u) - O_2 \Phi_2(t, hz; \Psi u) \right\|_Y . \tag{8}$$

Equation 8 requires well-defined solutions over the stated horizon and compatible output units. Zero discrepancy gives equality for the declared initial states, controls, times, and observation procedure under these maps. A small value is an explicitly indexed approximation. A computed maximum over sampled states, times, or inputs supplies a diagnostic; a uniform bound over an unsampled class requires further

Table 3: Comparison relations and their scope. The rows specify different preserved structures; they do not form a universal linear hierarchy.

Comparison	Preserved structure	Further information required
Coordinate conjugacy	Complete forward evolution under an invertible map and a common clock	Meaning-preserving readouts, control maps, and admissible state identification
Oriented orbits	Trajectory images and direction under a declared time change	Physical clock, durations, scheduled inputs, and exposure
Observable behaviour	Output paths or laws under specified preparations and controls	Hidden state, excluded interventions, and wider horizons
Approximate similarity	A declared error bound under fixed metric and quantifiers	Tolerance choice, coverage, and consequences of residual error
Attractor agreement	A specified asymptotic set or pattern	Transient paths, rates, basins, and finite-time consequences
Energy or action comparison	A specified functional, its stationary structure, or a scalar evaluated on a path	Mobility, variations, metric or noise, boundary terms, and measure
Boundary state sums	An identified boundary functional or operator	Interior insertions, normalization, and untested boundaries or refinements
Norm-relevant similarity	Legally or ethically material properties under a defended rule	Authority, evidence, exceptions, and contestability of the rule

analysis. The equality of output distributions in a stochastic model would require a comparison of laws, with the initial and noise distributions also specified.

Simulation and bisimulation in formal systems theory additionally require matching transitions while preserving observations; a bisimulation supplies matching in both directions. Approximate forms relax observational equality through a defined metric [11]. The numerical studies below do not implement a general bisimulation algorithm. Their exact claims follow from direct identities for the constructed models; their sampled metrics retain their limited domains.

An error threshold generally defines a similarity relation with different logical properties from equivalence. For example, scalar values 0, 0.75ε, and 1.5ε have successive distances below ε while their endpoint distance exceeds it. Approximate similarity at fixed tolerance is therefore generally nontransitive. Optimising over transformations can further change the meaning of the comparison. The admissible transformation class and the quantities it preserves should be specified before ranking models by their apparent similarity.

E.2 Variational, Boundary, and Normative Comparisons

This subsection locates energy, action, and boundary comparisons within the same account of preserved structure. Table 3 records the distinctions used throughout the comparative appendices.

An energy functional, an action functional, and an accumulated scalar value describe different objects. Equality of energies leaves mobility and dissipation unspecified. Equality of an action on one path leaves its value on other paths and its variational derivatives unresolved. Equality of full action functionals under a specified path map is a stronger statement whose consequences also depend on boundary conditions, admissible variations, and, for stochastic or quantum state sums, the measure and normalization.

Boundary equivalence compares complete boundary functionals or operators after internal variables have been integrated. It can hold while insertion observables in the interior distinguish the models. A renormalization or refinement-consistency claim must identify the coarse maps and the observables they carry. Such a claim for one finite family does not establish invariance across every complex or a continuum limit. These structural requirements are explicit in spin-foam coarse-graining work [3].

F Comparison Structures for Field Evolution

This section distinguishes mathematical relations between field evolutions through explicit constructions on a periodic ring. Each comparison specifies the field state space, observation map, temporal convention, and admissible transformation. Exact conclusions follow from direct operator identities. The numerical examples illustrate their consequences and verify the implementation; empirical applications require an additional interpretation of the field variables and observations.

F.1 State Space and Observation Metrics

This subsection fixes the domain and distance conventions used throughout the constructions. Let the spatial domain be the ring $\mathbb{T} = \mathbb{R}/(2\pi\mathbb{Z})$ and the state space be real $L^2(\mathbb{T})$. The spatial norm is normalised by the circumference. Two horizon-dependent comparisons are specified in Equation 9. The supremum convention controls the largest discrepancy over the interval, while the time-averaged convention measures accumulated squared discrepancy per unit time.

$$\begin{aligned} \|f\|_{\mathrm{rms}}^2 &= \frac{1}{2\pi}\int_{\mathbb{T}} |f(x)|^2\,dx, \\ d_{\infty,T}(u,v) &= \sup_{0\le t\le T} \|u(t)-v(t)\|_{\mathrm{rms}}, \\ d_{2,T}(u,v) &= \left[\frac{1}{T}\int_0^T \|u(t)-v(t)\|_{\mathrm{rms}}^2\,dt\right]^{1/2}. \end{aligned} \tag{9}$$

An observation operator can replace the full field in either comparison. For instance, the spatial mean discards every nonconstant Fourier mode. Equality of this observation defines classes of trajectories that may remain distinct in the field state space. Approximate similarity at tolerance ε requires both the metric and the horizon; it generally lacks transitivity. Constant fields with values 0, 0.75ε, and 1.5ε demonstrate this failure even before any observation reduction.

F.2 Fixed Spatial Coordinate Conjugacy

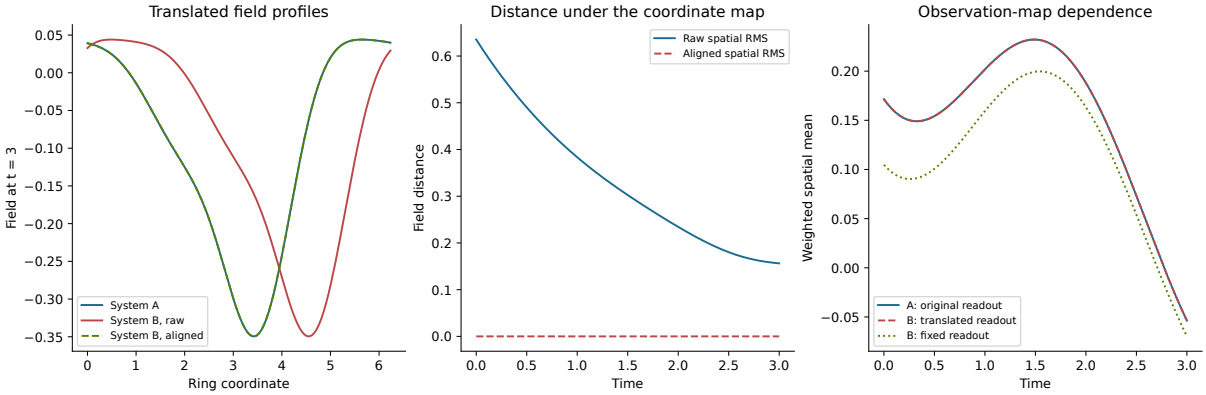
This subsection constructs a field equivalence under a fixed translation and examines the corresponding observation map. Define $(T_\delta f)(x) = f(x-\delta)$, an invertible isometry on the periodic state space. The autonomous part of the first system is $L_A = D\partial_x^2 - V_A(x)$ with domain $H^2(\mathbb{T})$, $D > 0$, and bounded positive potential. The second system uses $V_B = T_\delta V_A$. Spatial translation preserves the operator domain and commutes with the Laplacian. The resulting operator and semigroup identities are given in Equation 10.

$$L_B T_\delta = T_\delta L_A, \qquad S_B(t)T_\delta = T_\delta S_A(t), \qquad t \ge 0. \tag{10}$$

For controlled equations $\partial_t u_j = L_j u_j + b_j(x)g(t)$, the relation $u_B(t) = T_\delta u_A(t)$ holds for every common scalar control history g when both $u_B(0) = T_\delta u_A(0)$ and $b_B = T_\delta b_A$. A weighted spatial-mean observation $O_A(u) = \langle w_A u\rangle$ is preserved by the corresponding choice $w_B = T_\delta w_A$, because the ring measure is translation invariant. This preservation concerns transformed controls and observations; fixing a spatially selective observation weight defines a different comparison.

The numerical example takes $D = 0.04$, $V_A(x) = 0.4 + 0.2\cos x$, $b_A(x) = 0.8 + 0.3\sin 2x$, $g(t) = 0.3\sin(1.7t)$, and $w_A(x) = 1 + 0.5\cos(x - 0.2)$. The initial field is $0.7\cos x + 0.25\sin 3x$. A translation by 23 of 128 grid cells gives $\delta = 1.129010$. Figure 4 shows a maximum sampled raw field distance of 0.635315, exactly zero aligned field distance, and a transformed-readout discrepancy below 5.56×10^{-17}. The fixed-readout discrepancy reaches 0.066752. These observations illustrate the exact identities in Equation 10; the numerical equality alone would be insufficient to establish an identity for arbitrary initial fields and controls.

Figure 4: Fixed spatial coordinate conjugacy with translated coefficients, forcing profiles, initial fields, and observation weights. Raw field distance remains positive while the specified alignment preserves the complete field evolution and transformed observation.



F.3 Autonomous Time Reparameterisation

This subsection compares the temporal ordering of states with their physical timing. For $L = D\partial_x^2 - \lambda$ with $D, \lambda > 0$, define $L_B = cL_A$ for constant $c > 0$. The heat-decay semigroups satisfy Equation 11 for every initial field. Consequently the complete oriented forward orbits coincide under a positive rescaling of time.

$$S_B(t)u_0 = S_A(ct)u_0, \qquad t \geq 0. \tag{11}$$

The construction concerns forward semigroups, with no assumption of well-posed backward heat evolution. It establishes equality after the declared clock change and leaves same-time observations free to differ. Externally timed interventions, deadlines, and accumulated durations would require their own transformation rules. The identity makes no assertion about the existence or absence of additional state-space conjugacies.

Figure 5 uses $D = 0.12$, $\lambda = 0.3$, $c = 2.4$, and initial field $\cos x + 0.45 \sin 2x + 0.2 \cos 3x$. Its maximum sampled same-time RMS discrepancy is 0.236746, while the discrepancy after exact time alignment is zero. At physical time $t = 1$, the discrepancy is 0.230174. The plotted two-mode projection is a visualisation of the analytically established orbit identity, with the full third mode retained in all saved field arrays.

F.4 Observed Trajectories and Generator Identification

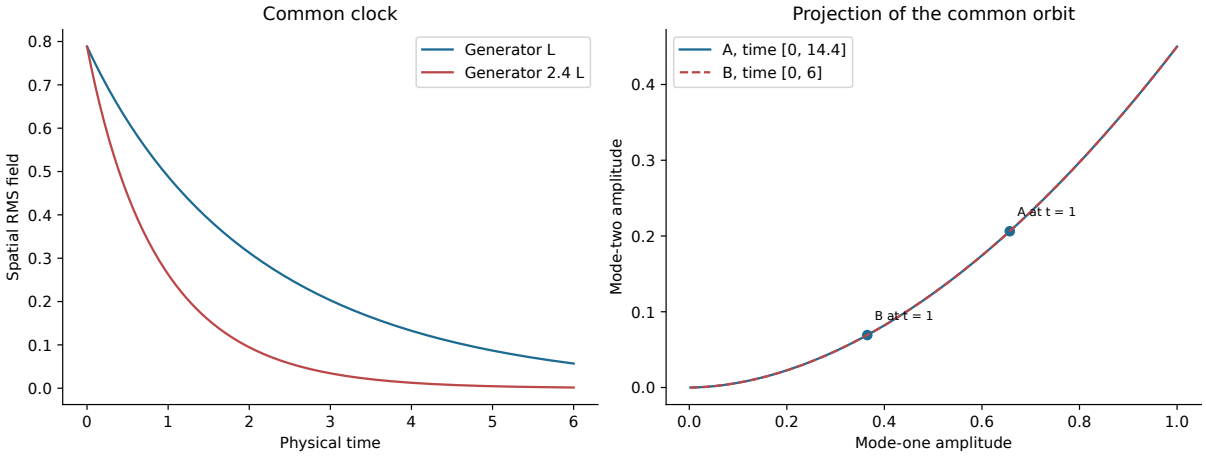
This subsection constructs complete trajectory agreement under one initial condition and separation under a declared probe. A Fourier mode $\cos kx$ under the heat-decay generator has rate $\mu_k = Dk^2 + \lambda$. The parameter pairs $(D_A, \lambda_A) = (0.1, 0.6)$ and $(D_B, \lambda_B) = (0.5, 0.2)$ give identical mode-one rate 0.7. Their restriction to the mode-one invariant subspace therefore agrees exactly, including every linear combination of $\sin x$ and $\cos x$. Complete noiseless observation of a mode-one trajectory does not determine the two parameters.

A reset to the common initial field $\cos 2x$ excites rates $\mu_{2,A} = 1.0$ and $\mu_{2,B} = 2.2$. The continuous-time probe discrepancy and its maximising time are given in Equation 12. Both quantities follow directly from the modal solution and the normalised spatial norm.

$$\begin{aligned} d_{\text{probe}}(t) &= \frac{e^{-t} - e^{-2.2t}}{\sqrt{2}}, \\ t_* &= \frac{\log(2.2)}{1.2} = 0.657048\ldots, \qquad d_{\text{probe}}(t_*) = 0.199936\ldots. \end{aligned} \tag{12}$$

Figure 6 distinguishes the baseline agreement from the probe separation. The full field observation is complete with respect to space and time for the baseline initial condition; it remains restricted with

Figure 5: Autonomous field evolutions with generators L and $2.4L$. The oriented forward orbit is shared, while observations at a common physical time differ. The phase plot displays a two-mode projection of the complete field trajectory.



respect to the set of initial conditions. This distinction supplies an explicit counterexample to identifying the generator from one agreeing trajectory.

Under the additional assumption that the generator belongs to this known two-parameter family, rates for two distinct modes identify the parameters. For modes one and two, $D = (\mu_2 - \mu_1)/3$ and $\lambda = (4\mu_1 - \mu_2)/3$. The audit recovers the chosen parameters from projections of saved field arrays to within 8×10^{-15}. This conditional identification result does not establish unrestricted mechanism identification from finite empirical observations.

F.5 Horizon, Readout, and Admissible Translation Classes

This subsection quantifies approximate similarity under distinct comparison conventions. Consider the common initial field $0.2 + \cos x$, decay $\lambda = 0.008$, diffusion $D = 0.002$, and a relative drift $\omega = 0.12$. The first field solves the heat-decay equation; the second has the additional transport term $-\omega \partial_x u$. Their solutions and raw spatial discrepancy are specified in Equation 13, with $a = D + \lambda = 0.01$.

$$\begin{aligned} u_A(x,t) &= 0.2e^{-\lambda t} + e^{-at}\cos x, \\ u_B(x,t) &= 0.2e^{-\lambda t} + e^{-at}\cos(x - \omega t), \\ \|u_A(t) - u_B(t)\|_{\mathrm{rms}} &= \sqrt{2}e^{-at}|\sin(\omega t/2)|. \end{aligned} \tag{13}$$

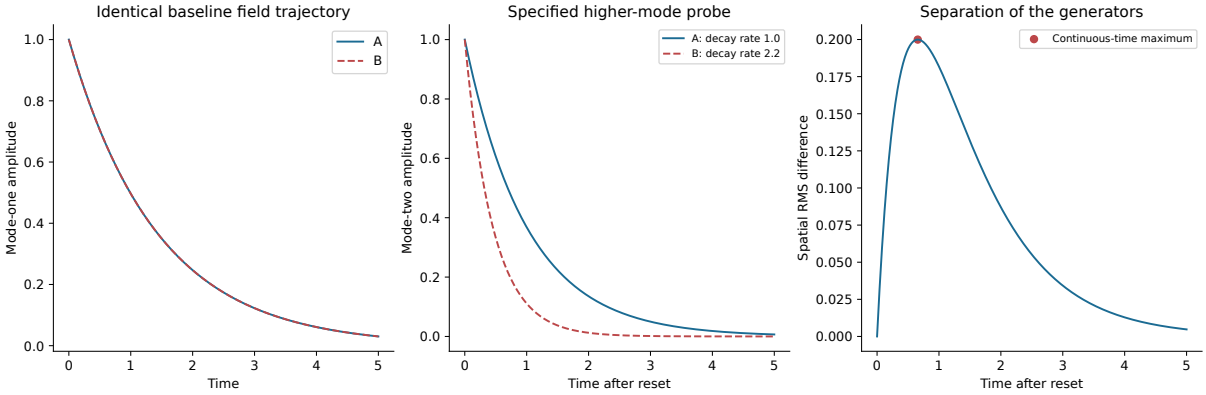
The spatial means agree at every time, while the full spatial observations differ. For the declared raw supremum metric, the distance increases from 0.042208 at horizon 0.5 to 0.165946 at horizon 2. Tolerance 0.1 therefore accepts the first comparison interval and rejects the second. The continuous supremum over every longer horizon is computed from the analytic expression, including its first and largest peak at $t = 23.427461$.

The effect of an admissible fixed translation is evaluated with the time-averaged metric. Write $I_0 = \int_0^T e^{-2at}\,dt$ and $I_\omega = \int_0^T e^{-2at}e^{i\omega t}\,dt$. For the convention $T_\delta f(x) = f(x - \delta)$, the optimal shift is $\delta_* = -\arg I_\omega$. Equation 14 gives the resulting minimum and the corresponding unaligned cost.

$$\begin{aligned} \inf_{\delta \in \mathbb{T}} d_{2,T}(u_A, T_\delta u_B)^2 &= \frac{I_0 - |I_\omega|}{T}, \\ d_{2,T}(u_A, u_B)^2 &= \frac{I_0 - \operatorname{Re} I_\omega}{T}. \end{aligned} \tag{14}$$

At horizon 20, the raw space-time RMS is 0.734627 and the optimal single-translation RMS is 0.427173. A time-dependent translation $\delta(t) = -\omega t$ produces zero aligned distance throughout the interval. This

Figure 6: Trajectory agreement on the excited mode-one invariant subspace and separation by a mode-two initial-condition probe. The continuous-time maximum is derived analytically and is distinguished from the maximum on the saved time grid.



time-dependent map introduces a moving frame and belongs to a larger comparison class than fixed coordinate conjugacy. Treating drift as removable can therefore discard transport information of substantive interest. Figure 7 separates the horizon effect, observation reduction, and the two translation classes; its left and right panels use the distinct metrics defined in Equation 9.

F.6 Attractors and Accumulated Observations

This subsection compares asymptotic agreement with path-sensitive quantities. Two heat-decay equations with $D = 0.05$ and decay rates $\lambda_A = 0.2$, $\lambda_B = 1$ have the same singleton zero global attractor on periodic L^2. The energy estimate in Equation 15 establishes exponential attraction of every bounded initial set. The derivative identity applies to regular solutions and extends by the usual semigroup approximation to the L^2 contraction bound.

$$\frac{1}{2}\frac{d}{dt}\|u(t)\|_{\mathrm{rms}}^2 = -D\|\partial_x u(t)\|_{\mathrm{rms}}^2 - \lambda\|u(t)\|_{\mathrm{rms}}^2, \qquad \|u(t)\|_{\mathrm{rms}} \le e^{-\lambda t}\|u(0)\|_{\mathrm{rms}}. \tag{15}$$

For the common initial field $1+0.4\cos x$, the spatial mean is $e^{-\lambda t}$. Its infinite-horizon integral is $1/\lambda$, giving 5 and 1, while the duration above a stipulated threshold $q = 0.4$ is $-\log(q)/\lambda$, giving 4.581454 and 0.916291. Figure 8 displays the transient and accumulated observations. Equal attracting sets therefore do not ensure equal histories, integrated exposures, or threshold durations.

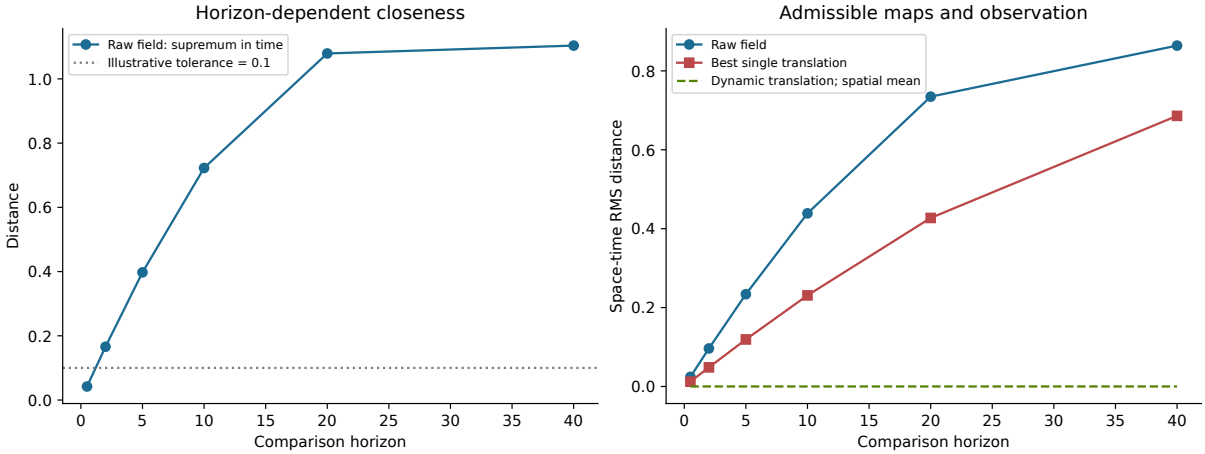
The exposure interpretation is illustrative. A connection to harm, opportunity, or obligation requires separately justified measurements and a normative rule. Attractor equality alone provides no such rule. The result instead identifies a mathematical distinction that a history-sensitive empirical or legal comparison may need to retain.

F.7 Numerical Evidence and Analytical Scope

This subsection records the evidential status and reproducibility of the computations. The heterogeneous translated system uses 128 periodic grid points, a centred second-order spatial Laplacian, and fourth-order Runge–Kutta time integration at step 0.0025. The remaining comparisons evaluate exact Fourier solutions. Their central identities and separating examples follow analytically; finite-grid observations provide illustrations and implementation checks.

Successive time-step refinement ratios are 16.5045 and 16.2502, while spatial consistency ratios are 3.9655, 3.9913, and 3.9978. The independent saved-output audit reproduces the fixed-translation costs by quadrature to 9.09×10^{-10} and verifies the modal rates, observation transformations, and contraction bounds. All eleven suite checks and fourteen audit checks pass. Source and configuration hashes, execution metadata, compressed field arrays, numerical tables, and both PNG and PDF figures are retained in `experiments/field_equivalence/`.

Figure 7: Approximate comparison of drifting field profiles under different horizons, observation maps, and admissible translations. The left panel uses a supremum-in-time field norm; the right panel uses time-averaged field RMS. Spatial-mean observation and time-dependent translation each produce zero distance for different reasons.



The scalar fields isolate comparison structures alongside the project's earlier three-field phenomenological model. They make no assertion that human and artificial systems instantiate the same equations. They establish specified identities and counterexamples, while leaving empirical field construction, admissible transformations, and norm-relevant observations to further substantive argument.

G Action, Mobility, and Field Comparison

This appendix develops action-based comparisons for a specified class of dissipative fields. Its method combines a variational specification with analytically controlled numerical examples, separating stationary states, trajectories, physical clocks, stochastic temporal laws, and scalar action values. The construction supplies mathematical support for the discussion of generative relations. Application to social or interpretive fields requires independent justification of the state variables, dynamics, observables, and admissible interventions.

G.1 Variational Structures and Comparison Data

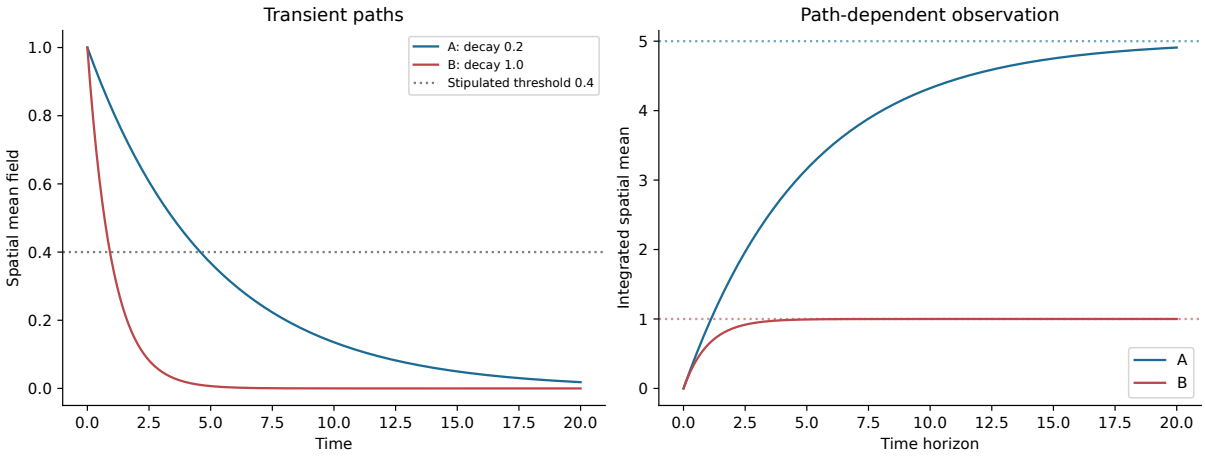
This subsection specifies the variational objects used in the comparison and their domains of interpretation. A classical stationary-action formulation selects histories by the vanishing first variation of an action under prescribed boundary variations. A gradient-flow formulation specifies an energy together with a dissipation mechanism; the latter converts energy derivatives into velocities [16, 19]. Consequently, an energy alone leaves temporal evolution underdetermined. The numerical construction below uses positive symmetric mobility and makes no claim that the programme's separate nonsymmetric, coupled three-field system admits this structure.

Let $\Omega = [0, 2\pi)$ be a dimensionless periodic ring. The real field is restricted to the invariant subspace $u(x,t) = a_1(t)e_1(x) + a_2(t)e_2(x)$, where $e_k(x) = \cos(kx)/\sqrt{\pi}$. The basis is orthonormal in $L^2(\Omega)$. The energy and its gradient flow are given in Equation 16.

$$E[u] = \frac{1}{2}\int_\Omega \left(0.15|\partial_x u|^2 + 0.4u^2\right) dx = \frac{1}{2}a^\mathsf{T} K a, \qquad \dot{a} = -MKa, \qquad K = \mathrm{diag}(0.55, 1). \quad (16)$$

The three mobilities are $M_1 = I$, $M_2 = 2I$, and $M_3 = \mathrm{diag}(0.4, 2)$. Each is positive. On the full ring, the third operator is defined as a Fourier multiplier with these eigenvalues on the selected modes and unit eigenvalues on remaining modes. It is therefore a specified nonlocal mobility. The chosen initial

Figure 8: Heat-decay fields with a common zero global attractor and distinct transient observations. The threshold and integrated spatial mean are explicitly stipulated path observables, without empirical or normative calibration.



condition has support only in the displayed modes, so their evolution is an exact invariant field restriction. This example does not depend on a spatial discretisation approximation to the evolution equation.

For a continuously differentiable trial history on $[0, T]$, define the residual functional in Equation 17.

$$I_M[a] = \frac{1}{4}\int_0^T (\dot{a} + MKa)^{\mathsf{T}} M^{-1} (\dot{a} + MKa)\, dt. \tag{17}$$

Equation 17 is a nonnegative least-squares residual once a drift and metric have been selected. Its zero set consists of the solutions of Equation 16 in the specified admissible class. Prescribing the initial field then fixes the deterministic solution. Generic stationary points of this path functional, with arbitrary endpoint constraints, need not have zero residual; those endpoint constraints may exclude every deterministic solution.

The same expression is the small-noise rate action for the finite-dimensional diffusion $da = -MKa\, dt + \sqrt{2\varepsilon M}\, dW$, conditional on its initial state. In this use, the rate controls leading logarithmic asymptotics of probabilities of path neighbourhoods as ε tends to zero. It does not supply a normalized finite-noise path density. Stochastic fluctuation-path formulations have established antecedents [15]; their relations to gradient flow require attention to the stochastic model and dissipation structure [1]. The stochastic claim here concerns a finite spectral system with constant nondegenerate diffusion. A continuum stochastic field theory would require a separately specified noise covariance and function-space analysis.

Expanding the square connects residual action to dissipation and endpoint energy, as shown in Equation 18.

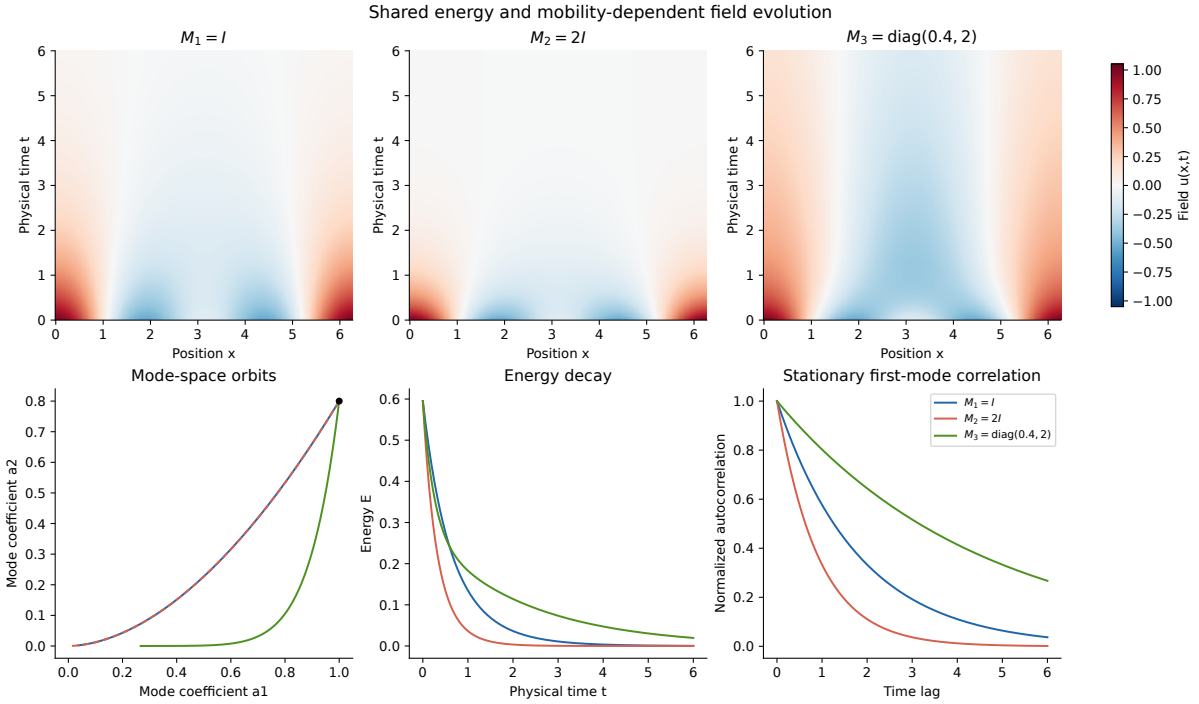
$$I_M[a] = \frac{1}{4}\int_0^T \left(\dot{a}^{\mathsf{T}} M^{-1} \dot{a} + (Ka)^{\mathsf{T}} MKa\right) dt + \frac{1}{2}\big(E[a(T)] - E[a(0)]\big). \tag{18}$$

Equation 18 follows directly from $\dot{a}^{\mathsf{T}} Ka = dE[a]/dt$. It distinguishes a time-integrated nonnegative dissipation expression from the endpoint contribution. For a deterministic relaxation, these terms cancel to give zero residual action. An isolated scalar integral therefore requires its functional definition and boundary data before comparison.

G.2 Shared Energy and Temporal Structure

This subsection compares the exact field evolutions generated by the three mobilities. All runs use $a(0) = (1, 0.8)$ and the physical interval $[0, 6]$. Exact mode amplitudes $a_k(t) = a_k(0)\exp(-m_k K_{kk} t)$ provide reference trajectories. Figure 9 presents their spatial histories, mode-space orbits, energies, and exact stochastic correlation functions.

Figure 9: Shared energy and mobility-dependent field evolution. The upper panels show exact spatial fields over a common physical interval. The lower panels compare geometric orbits under identity state identification, energy decay, and normalized stationary first-mode autocorrelation for the finite-dimensional diffusion extension. The reference and doubled-mobility orbits overlap geometrically; their physical clocks differ.



All three mobilities share the energy functional and its unique stationary field, $u = 0$. Along each solution, $dE/dt = -(Ka)^{\mathsf{T}} MKa \leq 0$. The equality $a^{(2)}(t) = a^{(1)}(2t)$ holds exactly, establishing orbit equivalence under identity state identification when a uniform rescaling of time is permitted. At the same physical time, their maximum field L^2 separation on the sampled interval is 0.307034. Thus physical-time trajectory equivalence imposes a stronger requirement than this particular orbit comparison.

The unequal-mode mobility changes the initial tangent direction: the sine of its angle with the reference tangent is 0.448998. Positive scalar time reparameterisation preserves tangent directions, so it cannot identify these two orbits through the common initial field under the identity state map. Their maximum sampled physical-time field distance is 0.329727. This statement fixes the physically interpreted field coordinates. Equivalence under an unrestricted change of state variables is a separate mathematical question and requires its own allowed-map class.

An explicit alternative identification clarifies this qualification. On the selected coefficient space, the homeomorphism $h(a_1, a_2) = (\mathrm{sgn}(a_1)|a_1|^{0.4}, \mathrm{sgn}(a_2)|a_2|^2)$ gives a time-preserving topological conjugacy between the M_1 and M_3 flows: $h(\Phi_1^t(a)) = \Phi_3^t(h(a))$. This identity follows by applying the indicated powers to the exponential mode solutions. The map changes the field interpretation and the prescribed initial-state correspondence; in particular, $h(1, 0.8) = (1, 0.64)$. It is a homeomorphism and need not be differentiable at zero. Thus identity-map orbit comparison and topological conjugacy supply different answers for the same pair of equations. Transporting physical observables would require a further explicit interpretation of h.

The diffusion extension illustrates an additional distinction. For each mobility its stationary law is Gaussian with covariance $\Sigma = \varepsilon K^{-1}$. This follows by direct substitution in the Lyapunov equation $(MK)\Sigma + \Sigma(MK)^{\mathsf{T}} = 2\varepsilon M$. With $\varepsilon = 0.1$, the diagonal variances are 0.181818 and 0.1. The normalized first-mode correlations at lag one are respectively 0.576950, 0.332871, and 0.802519. Equal stationary marginal distributions therefore coexist with unequal temporal laws at fixed physical time. These values are exact stochastic moments; stochastic sample paths were not simulated.

Table 4: Residual action for deterministic field histories under comparison generators. Rows identify the mobility generating the history; columns identify the mobility used in the residual and its metric. All histories begin at $(1, 0.8)$ and use $[0, 6]$.

Generating mobility	Action under M_1	Action under M_2	Action under M_3
M_1	0	0.07432806	0.10179066
M_2	0.07437521	0	0.21999995
M_3	0.09745974	0.20430053	0

G.3 Scalar Action and Historical Distinction

This subsection evaluates action as a comparison statistic and constructs histories occupying the same action level set. Table 4 first evaluates each deterministic history under each candidate generator, using Equation 17 with a common field identification, initial condition, and physical interval.

Every history has zero residual action under its own generator. Equality of these diagonal entries expresses successful evolution under three separately specified equations. It does not establish equality of those equations. The off-diagonal comparisons supply additional information about disagreement under a common state interpretation. Their asymmetry also shows that this cross-evaluation is a directed discrepancy, rather than a metric on models. A similarity analysis may use it only with an explicit reference generator, path ensemble, time interval, and normalization.

To test scalar action under one common functional, consider two trial paths on $[0, 4]$ with zero initial and final fields. Set $a^{[1]}(t) = (\sin(\omega t), 0)$ and $a^{[2]}(t) = (0, c\sin(\omega t))$, with $\omega = \pi/4$ and $c = \sqrt{(\omega^2 + 0.55^2)/(\omega^2 + 1)} = 0.754059$. Direct integration gives $I_{M_1}[a^{[1]}] = I_{M_1}[a^{[2]}] = 0.459675138$. Figure 10 displays their different spatial patterns and intermediate action accumulation. Their maximum field L^2 separation is 1.252440. These paths are admissible candidate fluctuations rather than zero-noise deterministic solutions. The construction demonstrates that a scalar action value, even with common endpoints and a common functional, leaves the intermediate history undetermined.

G.4 Coordinate Maps and Boundary Contributions

This subsection identifies structural information needed when action comparisons permit changes of representation. For an invertible constant linear coordinate map $v = Sa$, the transformed energy and mobility are given in Equation 19.

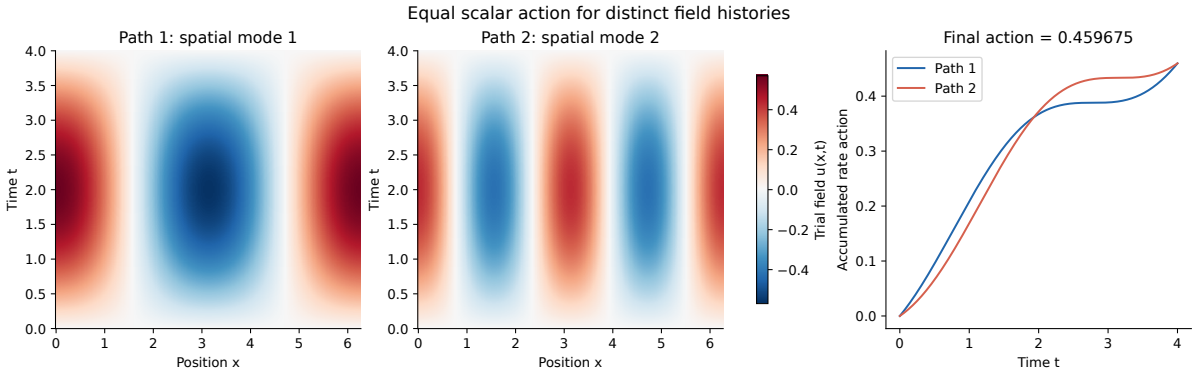
$$K_v = S^{-\mathsf{T}} K S^{-1}, \qquad M_v = S M S^{\mathsf{T}}, \qquad \dot{v} = -M_v K_v v, \qquad I_{M_v,K_v}[v] = I_{M,K}[a]. \tag{19}$$

Equation 19 follows from transforming both the drift and the quadratic residual metric. The diffusion covariance transforms to $2\varepsilon M_v$. Applying $S = \mathrm{diag}(2, 0.5)$ to a smooth two-mode test path gives original and transformed actions of 1.083807336. Keeping an identity residual metric after changing the coordinates gives 1.994733600. A coordinate transformation consequently preserves the comparison only when the associated metric and covariance are transformed. The calculation covers constant linear maps. Nonlinear Itô coordinate changes require the corresponding second-derivative correction to the drift.

Classical stationary-action equivalence involves a different comparison. Adding a total derivative to a Lagrangian shifts its action by $F(a(T), T) - F(a(0), 0)$ and preserves the bulk Euler–Lagrange equations under fixed endpoint variations [19]. If endpoint variations are allowed, boundary conditions require renewed analysis. Numeric equality of action values is therefore unnecessary for this particular equivalence of bulk variational equations. Conversely, the equal-action histories in Figure 10 show that numeric equality is insufficient for history identity. Comparison of action formulations should accordingly specify the complete functional, admissible histories, boundary data, transformations, and intended observables. A stochastic or quantum extension additionally requires its measure and interpretation of weights.

A direct calculation also qualifies the information supplied by a complete rate functional restricted to a boundary class. For $k > 0$, define $I_+[a] = \frac{1}{4}\int_0^T (\dot{a} + ka)^2\, dt$ for drift $-ka$ and $I_-[a] = \frac{1}{4}\int_0^T (\dot{a} - ka)^2\, dt$

Figure 10: Distinct field histories with equal final scalar action. Both trial fields begin and end at zero on $[0, 4]$ and are evaluated under the reference mobility. Their spatial modes differ. An analytically selected amplitude matches their total rate action while allowing distinct intermediate action accumulation and field histories.



for drift $+ka$, both with noise amplitude $\sqrt{2\varepsilon}$. Expansion gives $I_+[a] - I_-[a] = \frac{k}{2}\big(a(T)^2 - a(0)^2\big)$. Consequently, the functionals agree on every continuously differentiable zero-endpoint loop, although the deterministic drifts have opposite stability. Equality on this restricted path domain therefore leaves the forward generator undetermined. Initial-value evolution from other states or path classes with varied endpoints supplies additional discriminating information. This example is an analytic boundary-term calculation; it introduces no additional numerical run.

G.5 Verification and Scope of Interpretation

This subsection documents numerical verification and the scope of the illustrative conclusions. The protocol preceded the numerical execution. Main trajectories use exact exponentials; analytic time derivatives and integrals independently check residual-action quadrature. The maximum cross-action quadrature error is 3.55×10^{-7}. Periodic-grid basis orthonormality has maximum error 8.34×10^{-17}. Energy balance, the uniform clock map, equal-action construction, coordinate invariance, and the covariance Lyapunov identity agree at floating-point precision. An independent RK4 solver, with step sizes 0.2, 0.1, and 0.05, gives successive final-state error ratios from 16.54 to 18.04, approaching the fourth-order factor 16.

A separately timestamped post-run audit records the subsequently observed software environment and hashes of the preserved output files. It is a retrospective artifact audit, distinct from the execution-time record embedded in the numerical results. The analytic conjugacy and boundary-class qualifications above were added during integration without rerunning or replacing those outputs.

The simulation supports carefully scoped mathematical distinctions: a shared equilibrium or energy leaves mobility open; a shared orbit leaves its physical clock open; a shared stationary marginal leaves temporal correlations open; and a shared scalar action leaves the intermediate history open. A field similarity claim should state which of these structures it preserves, which transformations it admits, which histories or interventions it tests, and which norm and tolerance it uses. The examples introduce no new empirical claim about consciousness or normative standing. Their role in the discussion papers is to discipline the meaning of relational and historical comparison. Established variational and stochastic theories supply antecedents; conceptual priority is not asserted, and objections and corrections remain welcome. The paper's AI-use disclosure, licensing notice, and Responsible Use and Rights Reservation apply to this computational appendix.

H Finite-Group State Sums and Boundary Equivalence

This section constructs an exactly enumerable field model that separates boundary equivalence from interior statistics. It first identifies the representation-theoretic structure, then derives the boundary amplitude

and gluing rule, and finally compares two distinct microscopic models. The construction uses a classical positive-weight finite-group system. Its relation to spin foams is mathematical and explicitly limited.

H.1 Group Variables, Action, and Representation Labels

This subsection defines the degrees of freedom, measure, and action. It specifies the structural content that makes a dual spin-foam-type description available.

Take a strip of N square faces forming a disk. Each edge carries $g_e \in \mathbb{Z}_2 = \{-1, +1\}$, and each face has holonomy $U_f = \prod_{e \in \partial f} g_e$. Inversion leaves a group element unchanged. A vertex gauge transformation acts by $g_e \mapsto h_{s(e)} g_e h_{t(e)}^{-1}$ and preserves every face holonomy and every closed boundary holonomy. Open arcs can transform at their endpoints. Integration over an internal edge uses the normalized Haar average, with weight $1/2$ for each sign.

For $|t_f| < 1$, use the strictly positive face weights and dimensionless action in Equation 20.

$$\begin{aligned} w_{t_f}(U_f) &= 1 + t_f U_f = \frac{\exp(\kappa_f U_f)}{\cosh \kappa_f}, \qquad \kappa_f = \operatorname{arctanh} t_f, \\ S_{\boldsymbol{t}}[g] &= -\sum_f \log(1 + t_f U_f) = -\sum_f \kappa_f U_f + \sum_f \log\cosh \kappa_f. \end{aligned} \tag{20}$$

Equation 20 defines the statistical weight e^{-S} exactly. Each face weight has Haar average one. Dropping the local normalization factors changes the absolute boundary amplitude by a model-dependent multiplier. Although that multiplier cancels from conditional interior expectations within a fixed model, it cannot be silently removed when comparing absolute amplitudes between models.

The irreducible characters are $\chi_0(g) = 1$ and $\chi_1(g) = g$. Thus each face has character coefficients $1, t_f$. Integrating an internal edge annihilates terms whose two incident character labels sum to one modulo two. Surviving labels agree across that edge; this is the finite-group invariant-tensor constraint. The description therefore has actual representation labels and edge constraints, as in established finite-group spin-foam constructions [4, 10]. These labels are not geometric angular momenta of an SU(2) gravitational model.

H.2 Boundary Functionals and Composition

This subsection derives the complete boundary functional and identifies the effect of eliminating internal variables. The chosen normalization is retained throughout the calculation.

Let B be the product of the external boundary edges. The strip has $N-1$ internal edges. Its amplitude is given in Equation 21.

$$A_{\boldsymbol{t}}(B) = 2^{-(N-1)} \sum_{\{g_{\text{internal}}\}} \prod_{f=1}^{N} (1 + t_f U_f) = 1 + B \prod_{f=1}^{N} t_f. \tag{21}$$

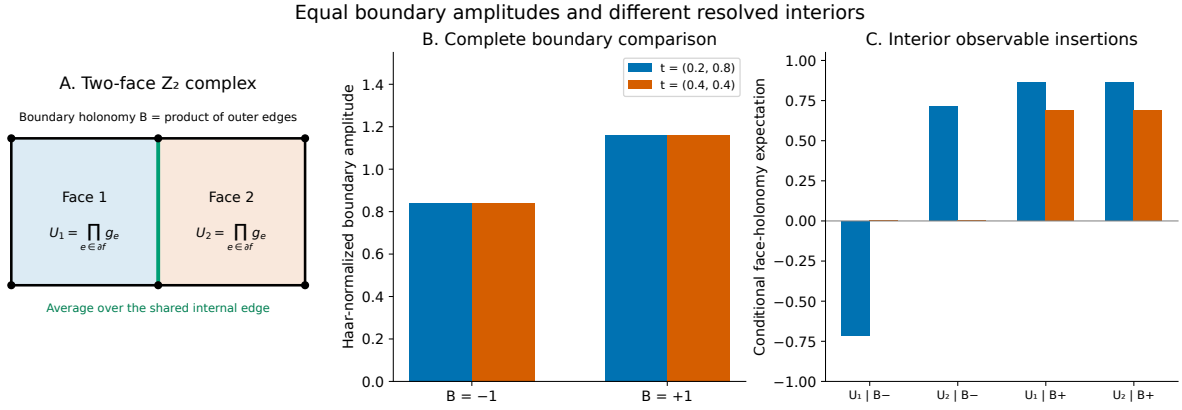
In Equation 21, the character constraints force all face labels on the connected strip to coincide. The all-zero label contributes one; the all-one label contributes the product of the couplings and the boundary holonomy. This establishes the formula for every strip length and coupling within the stated family. The subsequent finite enumerations check its implementation.

The kernel $K_t(a, b) = 1 + tab$ acts with Haar measure on boundary labels $a, b \in \mathbb{Z}_2$. When these labels describe open boundary arcs, their product $ab = B$ is gauge invariant. The individual arc labels retain their endpoint convention. Kernel composition obeys Equation 22.

$$\frac{1}{2} \sum_{b=\pm 1} K_s(a, b) K_t(b, c) = K_{st}(a, c), \qquad \operatorname{spec}(K_t) = \{1, t\}. \tag{22}$$

The spectrum in Equation 22 refers to the operator with normalized Haar integration, whose ordinary matrix is $K_t/2$. The characters diagonalize this operator. The multiplication rule also makes composition

Figure 11: A two-face finite-group complex with identical complete boundary amplitudes and different conditional interior observables. Haar averaging uses the declared normalized face weights. The boundary comparison retains both possible closed holonomies; the interior comparison retains the individual face identities.



associative. A coarse single-face parameter $t_{\text{eff}} = \prod_f t_f$ reproduces the fine strip's boundary amplitude. This is a restricted compatibility under elimination of internal edges. General cylindrical consistency would require specified maps and compatible measures across a wider family of complexes [3].

H.3 Identical Boundaries and Distinct Interior Laws

This subsection compares two microscopic coupling assignments and the observables retained or lost by their common boundary description. It reports both the exact analytical comparison and the executed enumeration.

The assignments $(t_1, t_2) = (0.2, 0.8)$ and $(0.4, 0.4)$ share $t_1 t_2 = 0.16$. Each therefore gives $A(+1) = 1.16$ and $A(-1) = 0.84$, and the same complete identified boundary operator. This is stronger than agreement in a single chosen boundary test. In this special normalized one-parameter boundary family, one sector value happens to determine the product; such identifiability does not extend automatically to a larger boundary space.

An interior insertion has conditional expectation $\mathbb{E}[U_1 \mid B] = (t_1 + B t_2)/(1 + B t_1 t_2)$. For $B = +1$, the two models give 0.8620689655 and 0.6896551724, a difference of 0.1724137931. For $B = -1$, the corresponding values are -0.7142857143 and 0. Their conditional interior laws therefore differ under the same face identification. Figure 11 displays the common boundary result and the distinct insertions.

The corresponding stronger comparison can be expressed by introducing sources for interior observables. Define a source-dependent amplitude and its first insertion through Equation 23.

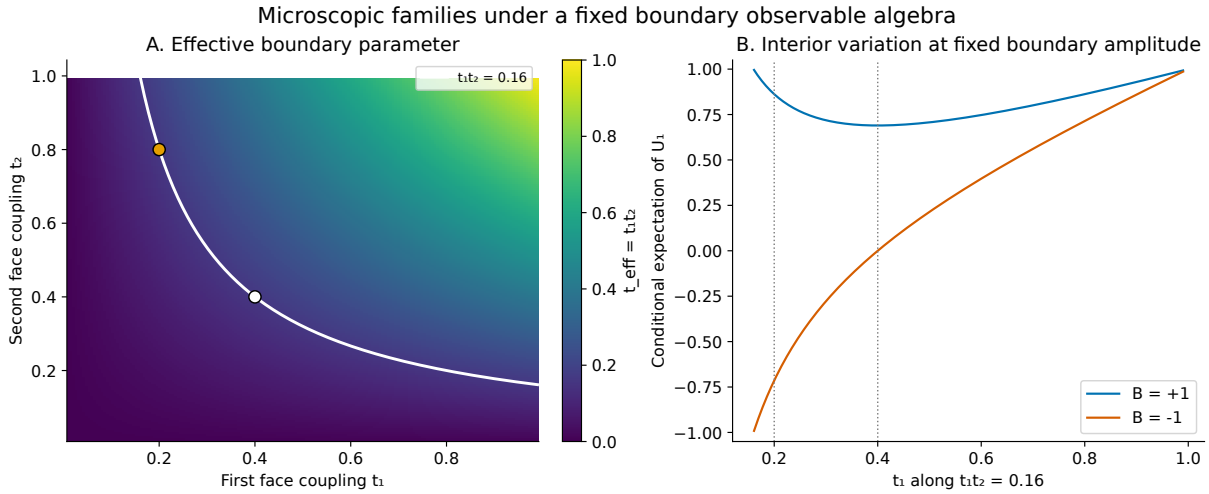
$$A_{\boldsymbol{t}}(B; J) = 2^{-(N-1)} \sum_{\{g_{\text{internal}}\}} \prod_f (1 + t_f U_f) \exp\left(\sum_f J_f U_f \right), \tag{23}$$

$$\frac{\partial}{\partial J_1} \log A_{\boldsymbol{t}}(B; J) \Big|_{J=0} = \mathbb{E}[U_1 \mid B].$$

Equation 23 makes the logical distinction explicit: equality at $J = 0$ leaves the derivatives with respect to independently specified interior sources undetermined. The selected models agree on their boundary functional while differing on this enlarged observable algebra. An effective description can preserve the retained observables without reconstructing the microscopic coupling factorization.

Figure 12 displays the continuous family $t_1 t_2 = 0.16$ within positive couplings. The boundary parameter is constant along that curve, while the conditional first-face expectation varies. Coarse boundary equivalence consequently relates a family of microscopic descriptions rather than a unique interior model.

Figure 12: Microscopic couplings with a common effective boundary parameter. The white curve in the left panel fixes $t_1 t_2 = 0.16$; the right panel evaluates an interior insertion along that curve. The family illustrates loss of resolved interior information under the specified boundary comparison.



H.4 Verification and the Scope of Spin-Foam Methods

This subsection records the finite checks and the interpretation permitted by the constructed model. It distinguishes the direct mathematical transfer from claims about gravitational dynamics.

Direct edge enumeration and dual character evaluation were compared for 48 strip conditions with one through eight faces, positive, signed, and zero couplings, and both boundary signs. Their maximum absolute discrepancy was 5.56×10^{-16}. All 64 two-face boundary-edge assignments agreed with the boundary-holonomy formula. Gauge invariance was exact across 8,192 edge-configuration and vertex-transformation combinations. Conditional expectations agreed with the analytic formula within 6.67×10^{-16}. Haar gluing and associativity were checked on 81 coupling pairs and 729 triples. The complete primary enumerations, action values, conditional probabilities, auxiliary checks, figure arrays, and source/output hashes are retained in `experiments/spinfoam_toy/`.

This finite construction uses a two-complex, group characters, invariant constraints, a boundary kernel, and a controlled sum over internal labels. These structures support precise comparisons between representations and resolutions. In gravitational spin-foam models, boundary states and amplitude composition require additional geometric and representation-theoretic data [17]. Even semiclassical connections to an action depend on the model and boundary regime: the EPRL four-simplex asymptotics studied by Barrett and colleagues contain multiple Regge-related contributions and prefactors [5]. A universal replacement of every amplitude by a single exponential of a chosen action would therefore omit material structure.

For the discussion papers, the transferable proposal is to make boundary data, hidden histories, admissible maps, and insertion observables explicit. A social interpretation of the group variables would require its own evidence and mapping. The present positive weights define a classical statistical model; they supply no oscillatory quantum amplitudes, gravitational geometry, or explanation of consciousness. The demonstrated boundary and interior distinction remains mathematically informative within that deliberately limited scope.

I Interpretation and Computational Records

This appendix records the scope and reproducibility of the shared mathematical studies.

I.1 Relational Interpretation and Issue-Relevant Equivalence

This section connects the mathematical comparisons to the companion discussion papers. It identifies the proposed analytical use and the substantive premises that the formal results leave open.

For Paper I, a reality comparison should identify which observations, relations, or interventions justify treating two entities alike. An equal recorded history need not identify its generator, whereas a coordinate difference can arise between descriptions that preserve the complete dynamics. Both distinctions matter: different appearance can accompany an equivalent description, and matching appearance can accompany different mechanisms. The appropriate conclusion depends on the declared comparison, rather than the word equivalence alone.

The action and finite-group calculations extend this argument to history-sensitive descriptions. An action can organise possible paths, and a boundary state sum can aggregate unresolved interiors. Their availability does not make every retained difference empirically accessible or relevant to a reality judgment. Conversely, equality of the selected aggregate cannot settle all questions about those interiors. The original three-field memory experiment remains a distinct example with its own state reset and context gate; the new studies examine equivalence between model descriptions more broadly.

For Paper II, an admissible comparison map must preserve the roles, time scales, options, and evidential distinctions material to the governing issue. A clock rescaling that removes a period of deprivation can destroy legal relevance while preserving an orbit. A boundary record that retains only a final agreement can erase the history through which agreement became available. The field representation should therefore make its observation map and omitted information inspectable. The legitimacy of a legal classification remains dependent on a norm and its institutional authority.

The numerical examples are constructed counterexamples and exact model identities. They are independent of the earlier 24-agent pilot and have no fitted relationship to its language outputs. They do not identify social dynamics as gravitational fields, establish a quantum mechanism of interpretation, or demonstrate consciousness. Their contribution is a disciplined set of comparison choices that can be tested and criticised before empirical or normative application.

I.2 Computational Records and Verification

This section identifies the executable record and the distinction between analytic claims and numerical checks. The field-equivalence, action, and finite-group experiments each preserve a local pre-run protocol, source code, arrays or exact enumerations, software metadata, results, and exported PNG/PDF figures. They use deterministic calculations and no external agent responses or fitted parameters.

The directories are `experiments/field_equivalence/`, `experiments/field_action/`, and `experiments/spinfoam_toy/`. The programme index in `experiments/field_comparison/README.md` links their reports and figures. Numerical refinements check discretized evolution; exact Fourier and finite-group calculations check the corresponding closed expressions. Finite tests supplement the analytic arguments and leave their stated function spaces, controls, and normalization intact. Source ledgers record the internet verification completed before citations were inserted.

References

[1] Stefan Adams, Nicolas Dirr, Mark A. Peletier, and Johannes Zimmer. Large deviations and gradient flows, 2012. URL https://arxiv.org/abs/1201.4601.

[2] Shun-ichi Amari. Dynamics of pattern formation in lateral-inhibition type neural fields. *Biological Cybernetics*, 27:77–87, 1977. doi: 10.1007/BF00337259.

[3] Benjamin Bahr. On background-independent renormalization of spin foam models, 2014. URL https://arxiv.org/abs/1407.7746. Version 2, 9 September 2014.

[4] Benjamin Bahr, Bianca Dittrich, Frank Hellmann, and Wojciech Kamiński. Holonomy spin foam models: Definition and coarse graining. *Physical Review D*, 87:044048, 2013. doi: 10.1103/PhysRevD.87.044048. URL https://doi.org/10.1103/PhysRevD.87.044048.

[5] John W. Barrett, R. J. Dowdall, Winston J. Fairbairn, Henrique Gomes, and Frank Hellmann. Asymptotic analysis of the EPRL four-simplex amplitude. *Journal of Mathematical Physics*, 50: 112504, 2009. doi: 10.1063/1.3244218. URL https://arxiv.org/abs/0902.1170.

[6] Pierre Bourdieu and Loïc J. D. Wacquant. *An Invitation to Reflexive Sociology*. University of Chicago Press, 1992. URL https://press.uchicago.edu/ucp/books/book/chicago/I/bo3649674.

[7] Andy Clark and David J. Chalmers. The extended mind. *Analysis*, 58(1):7–19, 1998. URL https://consc.net/papers/extended.html.

[8] Mark Coeckelbergh. Robot rights? towards a social-relational justification of moral consideration. *Ethics and Information Technology*, 12:209–221, 2010. doi: 10.1007/s10676-010-9235-5.

[9] Hanne De Jaegher and Ezequiel Di Paolo. Participatory sense-making: An enactive approach to social cognition. *Phenomenology and the Cognitive Sciences*, 6:485–507, 2007. doi: 10.1007/s11097-007-9076-9.

[10] Bianca Dittrich, Frank C. Eckert, and Mercedes Martin-Benito. Coarse graining methods for spin net and spin foam models. *New Journal of Physics*, 14:035008, 2012. doi: 10.1088/1367-2630/14/3/035008. URL https://arxiv.org/abs/1109.4927.

[11] Antoine Girard and George J. Pappas. Approximation metrics for discrete and continuous systems. *IEEE Transactions on Automatic Control*, 52(5):782–798, 2007. doi: 10.1109/TAC.2007.895849. URL https://www.georgejpappas.org/wp-content/uploads/2024/04/GP07-TAC.pdf.

[12] Ian Hacking. Making up people. *London Review of Books*, 28(16), 2006. URL https://www.lrb.co.uk/the-paper/v28/n16/ian-hacking/making-up-people. 17 August 2006.

[13] Pedro Henrique Luz de Araujo, Paul Röttger, Dirk Hovy, and Benjamin Roth. Principled personas: Defining and measuring the intended effects of persona prompting on task performance. In *Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing*, pages 26857–26886. Association for Computational Linguistics, 2025. doi: 10.18653/v1/2025.emnlp-main.1364.

[14] Michela Massimi. *Perspectival Realism*. Oxford University Press, 2022. doi: 10.1093/oso/9780197555620.001.0001.

[15] Lars Onsager and Stefan Machlup. Fluctuations and irreversible processes. *Physical Review*, 91: 1505–1512, 1953. doi: 10.1103/PhysRev.91.1505. URL https://journals.aps.org/pr/abstract/10.1103/PhysRev.91.1505.

[16] Mark A. Peletier. Variational modelling: Energies, gradient flows, and large deviations, 2014. URL https://arxiv.org/abs/1402.1990.

[17] Alejandro Perez. The spin-foam approach to quantum gravity. *Living Reviews in Relativity*, 16:3, 2013. doi: 10.12942/lrr-2013-3. URL https://link.springer.com/article/10.12942/lrr-2013-3.

[18] Alberto Romele. From relational personhood to genealogies of illusion: Media as the site of AI personification. In *Oxford Intersections: AI in Society*. Oxford University Press, 2026. doi: 10.1093/9780198945215.003.0172. Published online 8 August 2026; abstract-level verification.

[19] David Tong. Classical dynamics, 2004. URL https://www.damtp.cam.ac.uk/user/tong/dynamics/clas.pdf. University of Cambridge Part II Mathematical Tripos lecture notes, Michaelmas 2004 and 2005.

[20] A. M. Turing. Computing machinery and intelligence. *Mind*, LIX(236):433–460, 1950. doi: 10.1093/mind/LIX.236.433.

[21] J. C. Willems. Paradigms and puzzles in the theory of dynamical systems. *IEEE Transactions on Automatic Control*, 36(3):259–294, 1991. doi: 10.1109/9.73561. URL https://research.rug.nl/en/publications/paradigms-and-puzzles-in-the-theory-of-dynamical-systems/.

[22] Hugh R. Wilson and Jack D. Cowan. Excitatory and inhibitory interactions in localized populations of model neurons. *Biophysical Journal*, 12:1–24, 1972. URL https://isn.ucsd.edu/courses/beng260/reading/Wilson_Cowan_1972.pdf.

[23] Mingqian Zheng, Jiaxin Pei, Lajanugen Logeswaran, Moontae Lee, and David Jurgens. When "a helpful assistant" is not really helpful: Personas in system prompts do not improve performances of large language models. In *Findings of the Association for Computational Linguistics: EMNLP 2024*, pages 15126–15154. Association for Computational Linguistics, 2024. doi: 10.18653/v1/2024.findings-emnlp.888.