Cross-Substrate Phenomenal Comparison

A Typed Transformation-Transport Framework under Existential Uncertainty

Hongju Liu

24 September 2026

Affiliation: Independent researcher, Shenzhen, China
Report number: TA-TR-2026-15
Version: 1.0
Article type: Theoretical and methodological preprint
DOI: 10.5281/zenodo.22934654
Status: Not peer reviewed.

Authorship and accountability. Hongju Liu is the human author of record and responsible depositor. The paper was developed with substantial ChatGPT assistance in research, formalization, drafting, revision, coding, and publication preparation. AI-assisted review is not independent peer review.

Abstract

Comparing human, non-human animal, and artificial experience requires distinguishing whether a target has phenomenal experience from which aspects of its experience would be comparable if it does. Multidimensional profiles, behavioral similarities, and mechanistic correspondences provide important evidence, but do not by themselves settle measurement semantics, information loss, or the scope of cross-system inference. This paper proposes a typed transformation-transport framework whose basic comparison unit is a system-in-regime undergoing a specified intervention. Claims are distinguished by their targets: phenomenal existence, character, valence, perspectival organization, continuation, and token multiplicity. No universal human-centered phenomenal coordinate system is required, although each comparison requires an explicit local relational interface. The principal methodological object is a transport certificate separating observable causal correspondence, conditional phenomenal interpretation, and evidential updating about existence. Four propositions establish conditional non-identification under observational equivalence, a fiber criterion for query-preserving abstraction, guarded composition of set-valued constraints, and error propagation for metric interfaces. An exhaustively checked eight-state example shows that exact interventional correspondence can preserve one candidate phenomenal query while concealing a distinction relevant to another. A held-out intervention reveals the resulting model disagreement, but cannot empirically adjudicate it without an appropriate measurement connection. Bounded research protocols and a twelve-case conceptual regression suite operationalize the framework for human-animal-AI comparisons. The contribution is an integrated method for specifying and auditing transport claims, not a new sufficient condition for consciousness. In particular, the framework neither treats non-entailing evidence as worthless nor mistakes a formally valid conditional for an empirically established fact about experience.

Keywords: phenomenal consciousness; comparative consciousness; artificial consciousness; psychophysical bridges; causal abstraction; transportability; measurement invariance; partial identification

1. Introduction

How might human experience differ from the experience of a cat, a dog, or a future artificial agent? The question is not adequately answered by ranking intelligence, language ability, self-description, or computational complexity. It involves several distinguishable questions: whether experience occurs; what its contents and relational organization are; whether it has positive or negative valence; how it is organized into perspectives; and what, if anything, continues through changes in the implementing system.

Multidimensional animal-consciousness research has already challenged a single hierarchy of species, and the distinction between the distribution and quality of consciousness is well established [1,2]. Artificial-consciousness research likewise includes multilevel, multidimensional proposals [3]. The present paper does not claim to originate multidimensionality, non-anthropocentric comparison, interventionism, or the possibility of non-human forms of experience.

Its question is more specific: **under what conditions may a claim calibrated in one kind of system be transported to another kind of system, without transporting conclusions that the evidence does not support?** A correspondence in sensory discrimination may fail to preserve valence. Matching current contents may fail to establish a common perspective. Matching behavior or internal dynamics may increase the evidential support for consciousness without logically establishing its presence. A comparison therefore needs to retain the target of its inference, its measurement assumptions, and its unresolved premises.

The proposed unit of analysis is

$$(\text{system},\ \text{regime},\ \text{transformation},\ \text{evidence},\ \text{target claim}). \tag{1}$$

"System" does not mean a species label, software product, or account identifier. "Regime" includes the task, history, temporal scale, and relevant operating conditions. A transformation is specified by what is changed, how it is changed, and what is claimed to remain stable. The target claim identifies the kind of phenomenal question, rather than assuming that every successful comparison concerns consciousness as a whole.

The phrase **human experience as an epistemic anchor, not an ontological template** expresses two commitments. First, situated first-person judgments can contribute to calibration, without being treated as infallible public measurements. Second, the implementation details and conceptual vocabulary of humans need not define all possible forms of experience. Neither commitment licenses ignoring human measurement error, biological continuity, or evidence that differs substantially between animals and artificial systems.

A transformation-first approach is a research-design preference for changes that discriminate between explanations. It is not a claim that every intervention is superior to every observation. An intervention that disrupts the measurement channel or changes several mechanisms at once may be less informative than a well-characterized observational comparison. Static profiles remain useful descriptions; transformations help test what those descriptions retain and where their cross-system interpretation fails.

This paper contributes a concrete transport-certificate interface, explicit conditions for its interpretation and composition, and a reproducible formal example. The motivating human-animal-AI

question remains central. Digital copying and subject counting appear only as boundary cases that demonstrate why a result about one claim type must not silently become a result about another.

2. Related Work and the Scope of the Contribution

2.1 Multidimensional comparative consciousness

Birch, Schnell, and Clayton propose multidimensional consciousness profiles rather than a single interspecies scale [1]. Dung and Newen distinguish questions about consciousness distribution from questions about its quality, developing a species-sensitive framework [2]. Evers and colleagues propose a composite, multilevel, multidimensional heuristic for artificial consciousness [3]. These approaches establish important foundations for the present work. They should not be caricatured as necessarily static, anthropocentric, or incapable of representing heterogeneous systems.

The additional task addressed here is to specify the conditions under which a particular profile component or relational judgment can be carried from one system to another. A multidimensional description and a justified transport claim are different achievements.

2.2 Interventions and phenomenal structure

Klein and Barron use interventionist explanation, second-order interventions, and invariants to examine brain-experience relations [4]. Kleiner and Ludwig use variations to investigate which mathematical structures genuinely pertain to conscious experience, addressing the arbitrariness of simply assigning structures to experiential domains [5]. Consequently, neither transformation-based investigation nor concern about arbitrary phenomenal coordinates is new.

The present framework organizes these concerns around heterogeneous source and target systems, conditional phenomenal interpretation, and the composition of local comparisons. Its elementary mathematical results are not offered as replacements for the richer theories of structure or intervention developed in those works.

2.3 Animal tests and machine consciousness

Dung argues that suitably constrained animal-consciousness tests can also support attributions of machine consciousness, with particular attention to ad hoc test-passing, masking, and double dissociation [6]. His account is not a claim that such tests yield infallible deductive proof. It also recognizes limitations associated with experiences that differ substantially from familiar human cases.

Accordingly, "a test result does not logically prove consciousness" is not a sufficient novelty claim against that work. The present proposal instead supplies a record in which evidence about existence, a conditional interpretation of character, and a limitation on valence or continuation can be accepted or rejected separately. This is compatible with using tests as defeasible evidence.

2.4 Human-AI resemblance

Wang and colleagues require resemblance claims to identify the human reference class, AI system and version, task, property, measurement, perturbations, uncertainty, and permitted inference. They distinguish directional from symmetric measures, dynamical from static correspondence, and interventional from observational evidence [7]. These requirements substantially overlap with the methodological motivation of this paper.

The present manuscript focuses more narrowly on the semantics and composition of phenomenal claims whose phenomenal interpretation may retain an unresolved existence premise. It does not claim that explicit scope, directional comparison, or missing-value handling was absent from previous work. The comparison with [7] is restricted to its author-supplied abstract and verified bibliographic record; no claim is made about the absence of a particular formal result from its uninspected full text.

2.5 Causal transport, abstraction, and measurement

Transportability and heterogeneous-data integration have established formal treatments in causal inference [8]. Rubenstein and colleagues formalize consistency between causal models through agreement about interventions [9]; Beckers and Halpern analyze increasingly restrictive notions of causal abstraction [10]. More recently, Lorenz and Tull develop a compositional account of abstraction across models and queries [16]. The present paper therefore does not introduce causal transport, intervention-preserving abstraction, or compositional modeling.

Measurement invariance is likewise an established requirement for interpreting comparisons across populations [11]. The relationship between predicted experience and experience inferred from reports or behavior is a distinct issue in consciousness-theory testing [12]. Here, these concerns are applied to a specific interface problem: a causal mapping, a phenomenal interpretation, and a measurement model must not be treated as the same mapping.

2.6 Theory-dependent assessment

Theory-derived indicators already provide a way to investigate AI consciousness without waiting for agreement on a complete theory [13]. The present framework is compatible with that program. It permits evidence to alter support for consciousness, and does not require a deductive solution to other-minds problems before comparisons can be useful.

Its proposed contribution is an **integrated audit structure**: claim types, local interfaces, preserved premises, query-specific losses, and guarded network composition are represented together. The intended test of that contribution is whether the structure exposes otherwise hidden inferential errors and supports discriminating research designs. Neither a new name for the interface nor an elementary proof establishes historical priority. The literature comparison is focused rather than exhaustive.

3. Systems, Evidence, and Typed Claims

3.1 A system-in-regime

At a declared resolution, represent a system by

$$S_i = (X_i, \mathcal{I}_i, F_i, \mathcal{R}_i, \mathcal{H}_i), \tag{2}$$

where X_i is a space of physical or causal descriptions, $\mathcal{I}_i$ an intervention family, F_i conditional dynamics, $\mathcal{R}_i$ operating regimes, and $\mathcal{H}_i$ relevant history. States may be augmented with trajectories when the research question is not instantaneous. This representation does not establish that the system has a unique natural boundary.

For a biological comparison, the regime should identify the organism or population, developmental and learning conditions, task, and relevant sensory and motor channels. For an artificial system, it should identify the version, active mechanism, memory, tools, scheduler, and execution history

relevant to the claim. An external memory service is not automatically part of a phenomenal subject, but it cannot be excluded merely because it is accessed through an API.

3.2 Three levels that must remain distinguishable

The **causal-physical level** contains states $x_i \in X_i$, histories, mechanisms, and actual interventions. The **phenomenal level**, when experience occurs, contains experiential states or events $p_i \in \mathcal{P}_i$. The **evidence level** contains reports, judgments, behavior, neural measurements, and computational logs $o_i \in \mathcal{O}_i$.

A candidate psychophysical model Φ_i and measurement model M_i relate these levels. Where probabilistic modeling is appropriate, the observable likelihood may be written

$$\Pr(o_i \mid x_i, \iota_i, \Phi_i, M_i). \tag{3}$$

This notation does not impose a simple causal chain in which experience causes a judgment that causes a report. Physical inputs, task strategies, attention, and report mechanisms may contribute directly to observations. If measurement changes the system, the intervention must include that change. Defining experience through a report and then validating the definition against the same report does not provide an independent test [12].

3.3 Six claim types, not six universal consciousness scores

Let

$$K = \{E, Ch, Val, Persp, Cont, Mult\}. \tag{4}$$

Type	Target question	A neighboring quantity that does not answer it by definition
E: existence	Does experience occur in the specified system or interval?	Intelligence, language fluency, task accuracy
Ch: character	What contents or relational qualities occur or change?	Latent-vector dimension, classification accuracy
Val: valence	What positive, negative, mixed, or neutral phenomenal value occurs?	Reward number, loss function, avoidance
$Persp$: perspective	How are experiences organized into perspectives or wholes?	Software-module count, controller hierarchy
$Cont$: continuation	Which cross-time experiential or perspectival successor relations obtain?	Memory equality, checkpoint restoration
$Mult$: multiplicity	Which distinct occurrences or tokens obtain under a specified individuation rule?	Logical role count, repeated descriptions, sampling count

These are types of question, not six independent substances or a fixed six-dimensional experiential space. Valence may be an aspect of character. Perspectival organization may partly determine whole-level character. A theory may couple these fields, but then the coupling is part of the theory, not a consequence of using one undifferentiated word, "consciousness."

An existence proposition $E_i = 1$ is useful in some models. It does not make richness, intensity, organization, or moral importance binary. Models of borderline or graded existence can be

considered separately; the binary non-identification result below is stated only for model classes containing opposing binary claims.

3.4 A typed transformation signature

For an intervention $\iota_i : x_i \rightsquigarrow x_i'$, write

$$\Sigma_{\Phi_i}(\iota_i; x_i) = \langle \Delta E, \Delta Ch, \Delta Val, \Delta Persp, \Delta Cont, \Delta Mult \rangle. \tag{5}$$

Each field records the target property, predicted or inferred change, conditions, and evidential status. A field may be preserved, changed, created, terminated, split, merged, or unresolved, when those descriptions have defined semantics for that target.

Crucially, the following are different:

- `unknown`: the relevant claim has not been determined;
- `out-of-scope`: the model does not address the claim;
- `undefined-if-no-experience`: the phenomenal property lacks a referent under a no-experience model;
- `no-change-detected`: no change was detected at the stated measurement resolution;
- `conditionally-preserved`: preservation follows only under stated premises.

None is automatically equivalent to zero. These distinctions matter even when the eventual output is a simple table rather than a formal model.

4. From Causal Correspondence to a Transport Certificate

4.1 Causal correspondence

Let S be the source and T the target. Select a declared target subdomain $U \subseteq X_T$ and a matching or abstraction map

$$\alpha : U \longrightarrow X_S. \tag{6}$$

The map may be many-to-one. Its direction is target-to-source because it expresses which target distinctions are retained by the source description; this is compatible with transporting a calibrated source claim toward the target.

For matched interventions ι_T and ι_S, an exact interventional-consistency condition is

$$\alpha(F_T(x, \iota_T)) = F_S(\alpha(x), \iota_S), \qquad x \in U. \tag{7}$$

Both sides must be defined, including the relevant successor domain. In stochastic models the corresponding condition concerns intervention distributions, not sample-by-sample trajectory identity. Equation (7) uses established causal-abstraction ideas [9,10]; it is not a new consciousness law.

Exact correspondence is one possible comparison target, not a universal prerequisite for useful research. Approximate, directional, or difference-seeking comparisons can be specified instead. What matters is that the criterion be fixed and interpreted at its declared level.

Even exact correspondence establishes only the causal relation encoded by α and the tested interventions. A phenomenal interpretation additionally depends on Φ_S, Φ_T and the measurement models. Naming an uncalibrated target variable "pain" does not provide that connection.

4.2 No universal phenomenal coordinates, but a local shared interface

Comparison cannot dispense with every common relation and still claim to compare the same property. The appropriate alternative to a universal human template is therefore not complete absence of shared semantics, but **explicit local comparability**.

For a claim type k, an interface Z_k may consist of a partial order, discriminability relation, transition graph, finite categories, or a justified metric space. Conditional on experience and the relevant interpretation, maps may be defined as

$$a_S^k : \mathcal{P}_S \to Z_k, \qquad a_T^k : \mathcal{P}_T \to Z_k. \tag{8}$$

Nothing requires $\mathcal{P}_S = \mathcal{P}_T$, or either map to be invertible. One study may compare ordering under a sensory manipulation; another may compare temporal grouping. These local interfaces need not combine into a single universal coordinate system. Where endpoint interfaces differ, an explicitly calibrated relation between them can replace a shared Z_k.

A target distinction excluded by an interface is unrepresented, not thereby absent. Conversely, an unrepresented computational variable is not automatically a new phenomenal dimension. Mechanistic discovery and phenomenal interpretation remain separate claims.

4.3 Certificate structure

Define a transport certificate schematically by

$$C_{S\to T}^k = \langle k, \mathcal{U}, J, \alpha, Z_S^k, Z_T^k, R_k, \Gamma, M_S, M_T, \mathcal{E}, \mathcal{L}, V \rangle. \tag{9}$$

Here $\mathcal{U}$ identifies endpoint systems and validity domains; J pairs interventions; α records causal correspondence; Z_S^k, Z_T^k give interface semantics; R_k restricts permitted cross-interface relations; Γ contains applicability and interpretive premises; M_S, M_T are measurement models; $\mathcal{E}$ records evidence and dependencies; $\mathcal{L}$ records losses and uncovered distinctions; and V identifies the version.

A certificate is an auditable claim record, not a certification of truth. It must distinguish assumptions that have independent support, assumptions provisionally adopted, and assumptions for which no test is currently available. The three principal statuses are:

Causal correspondence. The record supports specified relations between mechanisms or observations, without asserting a target phenomenal interpretation.

Conditional phenomenal transport. Under the existence and bridge premises in Γ, a target phenomenal query obeys a specified constraint. Unresolved premises remain attached to the conclusion.

Existence-evidence update. Data change the relative support for models of experience and its absence through stated likelihoods or another explicit evidential argument. This is a separate inference, not an automatic promotion of a conditional's antecedent.

A useful certificate must restrict a query. If R_k allows every possible target output, it supplies no information about that query. If a conditional restriction has no supported measurement connection and no observable consequence, it is a conceptual model constraint, not an empirically validated transport result. This prevents completeness of documentation from being mistaken for scientific success.

4.4 Non-entailment is not evidential irrelevance

A structural match need not logically entail experience in order to support it. In a suitable model class it is entirely possible that

$$\Pr(E_T \mid \mathcal{E}) > \Pr(E_T). \tag{10}$$

The magnitude and direction of updating depend on the alternatives, likelihoods, and priors, not on the vocabulary of type safety. Familiar biological evidence and artificial-system evidence need not receive equal assessments. The present framework does not require every organism to pass an infallible test before attributing experience [6,13,14].

What must not happen is an unnoticed change from "if the target experiences, this bridge predicts a valence relation" to "the target experiences and has that valence." A model may justify both conclusions, but it must supply the additional grounds. Section 5 states a limited circumstance under which the available evidence cannot resolve existence; it does not assert that this circumstance holds universally.

5. Four Formal Conditions for Transport

The following propositions use elementary ideas from identification, factorization, relation composition, and metric error propagation. Their value here is to make the certificate's obligations explicit. They do not identify a mechanism that produces experience, and their underlying mathematics is not claimed as novel.

5.1 Conditional non-identification of existence

Let $\mathfrak{M}$ be a candidate psychophysical-measurement model class. Let $\mathcal{P}$ contain the complete experimental protocols available to the study, including any adaptive intervention and stopping rules. Model m induces a transcript distribution P_m^e for protocol e.

Proposition 1. Suppose $m_0, m_1 \in \mathfrak{M}$ satisfy

$$E_T(m_0) = 0, \qquad E_T(m_1) = 1, \tag{11}$$

while

$$P_{m_0}^e = P_{m_1}^e \qquad \text{for every } e \in \mathcal{P}. \tag{12}$$

No decision procedure restricted to those protocols can identify E_T almost surely correctly under both models.

Proof. Fix any permitted protocol and decision rule. If it returns 0 almost surely under m_0, equality of transcript distributions implies that it returns 0 almost surely under m_1. It therefore

cannot return 1 almost surely under m_1. Any external randomization can be included in the common transcript distribution. Because the equality holds for every permitted complete protocol, changing to another permitted protocol does not remove the obstruction. □

The proposition is conditional on the model class and protocol family. It does not establish that a phenomenally empty counterpart of every real system is physically possible. It does not imply that all existence theories are observationally equivalent. A new protocol with different predicted evidence distributions invalidates equation (12) for the enlarged protocol family.

Matching a source system's transformations may leave a pair satisfying equations (11)-(12) in the candidate class. It may instead exclude some no-experience models. The empirical question is which situation the justified model class and measurements support.

5.2 Query-preserving abstraction: the fiber criterion

Fix a candidate model, a validity domain, and a well-defined target query. Let $\alpha : U \to X_S$ and let

$$f_T^k : U \to Y_k$$

be the target-model query for type k. A query may describe a state property or, by including intervention and history in its input, a transformation signature. Y_k need not be numeric.

Proposition 2. A function $h : \alpha(U) \to Y_k$ exists such that

$$f_T^k = h \circ \alpha \tag{13}$$

if and only if f_T^k is constant on every fiber of α:

$$\alpha(x) = \alpha(x') \implies f_T^k(x) = f_T^k(x'). \tag{14}$$

When it exists, h is unique on $\alpha(U)$.

Proof. Equation (13) immediately implies equation (14). Conversely, for $z \in \alpha(U)$, choose any x satisfying $\alpha(x) = z$, and define $h(z) = f_T^k(x)$. Fiber constancy makes the definition independent of the representative. Surjectivity onto $\alpha(U)$ gives uniqueness. □

Thus, an irreversible map can preserve a particular query without preserving all target structure. The map is inadequate for that query if it collapses two states to which the query assigns different values. This is **query-specific abstraction aliasing**, not merely low similarity.

There is an additional calibration requirement. Proposition 2 establishes the existence of some h; it does not establish that h equals the source query f_S^k. Directly transporting the source judgment requires

$$h(z) = f_S^k(z), \qquad z \in \alpha(U), \tag{15}$$

with a shared interpretation of the codomain. Target abstraction sufficiency and source-target semantic calibration are separate obligations. This is why successful preservation of a character query cannot automatically authorize a valence query using the same state map.

The criterion is structural and model-relative. It is not an algorithm for discovering the actual phenomenal query from unlabelled mechanism data. Its failure can be witnessed in a formal model; empirical use additionally requires access to evidence about the relevant target distinction.

5.3 Guarded composition of set-valued constraints

Suppose three endpoint interfaces of the same claim type are Z_S, Z_T, Z_U. A source constraint $Q_S \subseteq Z_S$ contains the actual value z_S. Certificates provide relations

$$R_{ST} \subseteq Z_S \times Z_T, \qquad R_{TU} \subseteq Z_T \times Z_U.$$

For a relation R, define its image by

$$R[Q] = \{y : \exists x \in Q,\ (x, y) \in R\}. \tag{16}$$

Proposition 3. Assume the certificates have compatible claim types, interface meanings, regimes, temporal and historical indices, and existence premises. They refer to the same actual intermediate value z_T. Under their joint guard $\Gamma_{ST} \wedge \Gamma_{TU}$, suppose

$$(z_S, z_T) \in R_{ST}, \qquad (z_T, z_U) \in R_{TU}. \tag{17}$$

Then

$$z_U \in R_{TU}[R_{ST}[Q_S]]. \tag{18}$$

Proof. The source inclusion and first relation place z_T in $R_{ST}[Q_S]$. Applying the second relation gives equation (18). □

If additional, jointly valid constraints give $z_T \in Q_T$, one may replace the intermediate set by $R_{ST}[Q_S] \cap Q_T$. An empty intersection or empty image signals incompatible premises, mappings, domains, or measurements; it does not by itself diagnose absence of experience. At most, it rejects a conjunction of assumptions. Selecting existence as the failed assumption requires an additional argument.

Transport is therefore not intrinsically non-transitive. It is **not unconditionally transitive**. A human-to-animal character relation and an animal-to-AI continuation relation cannot be composed as a claim about AI valence. Two compatible relations about the same intermediate condition can compose.

Constraints from mutually exclusive candidate models must retain their model labels. Combining them as though they were jointly true can eliminate valid possibilities. Similarly, the same source dataset arriving through two network paths is still one dataset. Logical intersection of valid constraints is distinct from multiplication of evidence likelihoods. Statistical independence is not required for the set inclusion in Proposition 3 or for intersecting jointly valid constraints. If the sets are estimated confidence regions rather than guaranteed constraints, their joint coverage must be controlled; separate marginal coverage levels do not automatically carry over to the intersection.

5.4 Error propagation on a justified metric interface

For a particular type, suppose (Z_i, d_i) are justified metric interfaces, and $g_{ST} : Z_S \to Z_T$, $g_{TU} : Z_T \to Z_U$ are calibrated maps. In a common validity region, assume

$$d_T(z_T, g_{ST}(z_S)) \le \varepsilon_{ST}, \qquad d_U(z_U, g_{TU}(z_T)) \le \varepsilon_{TU}. \tag{19}$$

Assume also that g_{TU} is L_{TU}-Lipschitz over the relevant points.

Proposition 4. Under these assumptions,

$$d_U(z_U, g_{TU}(g_{ST}(z_S))) \le \varepsilon_{TU} + L_{TU}\varepsilon_{ST}. \tag{20}$$

Proof. Insert $g_{TU}(z_T)$, apply the triangle inequality, and then apply the Lipschitz condition and equation (19). □

The bound concerns one local interface. It is not a percentage of shared consciousness, nor a general distance between subjects. Error bounds must be justified by calibration, model assumptions, or validation; they must not be invented to make the record look quantitative. Where no defensible metric is available, the relational or set-valued representation is preferable. If the two inequalities in equation (19) hold with failure probabilities at most δ_{ST} and δ_{TU}, respectively, and the Lipschitz condition is guaranteed on the relevant region, equation (20) holds with probability at least $\max\{0, 1 - \delta_{ST} - \delta_{TU}\}$ by the union bound. This statement requires no independence. Uncertainty in the Lipschitz condition, data-dependent selection of the region, or selection among multiple paths requires its own coverage accounting.

5.5 What the conditions jointly permit

These propositions establish a disciplined positive procedure: identify a query and interface; test whether the abstraction preserves that query; calibrate the source-target interpretation; retain all premises; and compose only compatible relations, with uncertainty and evidence dependencies intact. A successful field need not confer success on every other field. This yields local knowledge without requiring either universal phenomenal equivalence or universal skepticism.

6. A Finite-State Illustration

6.1 Construction and interpretation

The following example is constructed for this paper. It is not a biological experiment, a simulation of a conscious agent, or evidence that Boolean variables feel anything. Its functions called "candidate character" and "candidate valence" are stipulated interpretations inside competing toy models.

Let the source state space be

$$X_S = \{(u, b) : u, b \in \{0, 1\}\},$$

and the target state space be

$$X_T = \{(a, b, h) : a, b, h \in \{0, 1\}\}.$$

Define

$$\alpha(a, b, h) = (a \oplus h, b), \tag{21}$$

where $\oplus$ is exclusive OR. Every source state has two preimages, so the projection discards a distinction involving h.

The source has an identity action I, a first-coordinate action $C_S(u, b) = (1 - u, b)$, and a second-coordinate action $V_S(u, b) = (u, 1 - b)$. Their target counterparts are

$$C_T(a, b, h) = (1 - a, b, h), \qquad V_T(a, b, h) = (a, 1 - b, h). \tag{22}$$

Equation (7) holds for all eight target states and all three matched actions I, C, V. The letter V names an action; it does not certify that this action changes phenomenal valence.

6.2 Agreement on the calibration regime

Suppose two candidate interpretations both posit target experience and share

$$f_T^{Ch}(a, b, h) = a \oplus h.$$

They disagree about a binary valence-category query:

$$f_{T,1}^{Val}(a, b, h) = b, \qquad f_{T,2}^{Val}(a, b, h) = b \oplus h. \tag{23}$$

The categories 0 and 1 are not magnitudes of suffering. If calibration only covers $h = 0$, the interpretations agree on all four calibration states and on every I, C, V successor of those states.

Over the full target space, f_T^{Ch} and $f_{T,1}^{Val}$ are constant on each fiber of α, but $f_{T,2}^{Val}$ is not. For example,

$$x = (0, 0, 0), \qquad x' = (1, 0, 1)$$

have the same projection $(0, 0)$, while $f_{T,2}^{Val}(x) = 0$ and $f_{T,2}^{Val}(x') = 1$. Under interpretation 2, the abstraction is insufficient for the valence query even though it preserves the candidate character query.

6.3 A held-out transformation

Introduce a target action excluded from calibration:

$$H_T(a, b, h) = (1 - a, b, 1 - h). \tag{24}$$

For every target state,

$$\alpha(H_T(x)) = \alpha(x). \tag{25}$$

Interpretation 1 predicts preserved valence category; interpretation 2 predicts a flipped category. Both predict preserved candidate character and the same projected output. The target action therefore witnesses a model-relative distinction concealed by the interface.

Query	Prediction under H_T	Status
Projected causal readout α	Preserved	Direct algebraic result
Candidate character	Preserved in both interpretations	Conditional model result
Candidate valence, interpretation 1	Preserved	Conditional model result
Candidate valence, interpretation 2	Flipped	Conditional model result
Actual target experience	Not determined	No empirical consciousness inference

6.4 Prediction divergence is not yet empirical discrimination

If all available measurements are $o = \alpha(x)$, the two interpretations still yield the same observable results. A mathematical disagreement about their candidate phenomenal outputs does not make those outputs measured.

An empirical test requires a measurement connection supported independently of naming the hidden bit, or observable consequences on which the competing models disagree. Adding a probe for $b \oplus h$ establishes access to that mechanism variable. It does not, by itself, establish that the probe measures unpleasantness. This is the distinction between an **abstractly discriminating transformation** and an **empirically discriminating protocol**.

A third candidate model may posit no target experience while preserving every stated physical transition and readout. If this model is admitted and the readouts exhaust the available evidence, Proposition 1 applies. The example does not establish that such a model is true or universally admissible.

6.5 Executed checks

The standard-library Python program in Appendix B was executed for this manuscript. It exhaustively checks the finite domains specified above and performs two supplementary checks of relation composition and a metric error bound.

Check	Scope	Result
Interventional commutation	8 states $\times$ 3 actions = 24 equalities	All satisfied
Projection invariance under H_T	8 states	All satisfied
Calibration-successor agreement	4 states $\times$ 3 actions = 12 comparisons	All agree
Preservation-versus-flip prediction under H_T	8 states	Divergence at every state
Fiber constancy of candidate character	Every fiber	Satisfied
Fiber constancy of candidate valence 1	Every fiber	Satisfied
Fiber constancy of candidate valence 2	Every fiber	Violated
Baseline replay ignoring action C	8 intervention checks	Distinguished in every case
Compatible and incompatible relational paths	One of each	Valid composition; empty image for incompatible path
Metric-composition bound	11 rational inputs	Bound 13/100 satisfied

These are algebraic checks, not 24 consciousness experiments. The replay check concerns the specifically defined device that ignores an intervention; it is not a general claim that every replay or lookup implementation fails every relevant causal test.

7. Comparing Humans, Cats, Dogs, and Artificial Agents

7.1 Unequal evidence does not require a universal hierarchy

The framework does not place all candidate systems in an identical state of evidential uncertainty. The 2024 New York Declaration on Animal Consciousness states that there is strong scientific support for attributing conscious experience to other mammals and birds [14]. This is an attributed field declaration, not a new result of this paper, and it does not establish a complete mapping between human, feline, and canine experiential qualities.

For artificial systems, this paper supplies no new empirical confirmation of experience in a particular model. Nor does lack of such confirmation establish absence. Existing theory-based assessment provides a basis for examining specific mechanisms and indicators [13]; the present method records what each comparison permits without updating a historical assessment into a claim about every system available today.

The table below is a research-task decomposition, not a measured species profile.

Target	Human anchor material	Animal-side investigation	Artificial-system investigation	Inference requiring an additional bridge
Character	Discrimination, similarity judgments, reports, intervention responses	Species-appropriate stimuli, behavior, and mechanism measurements	Mechanism-level representations and intervention responses	Same task performance → same experience
Valence	Affective judgments under controlled conditions	Converging affect-related evidence and alternative explanations	Reward, policy, regulation, and candidate valence mechanisms separately	Negative reward → suffering
Temporal character	Ordering, duration, and grouping judgments	Task-appropriate temporal responses and mechanisms	Logical steps, physical timing, memory, and environmental coupling	Clock frequency → subjective time
Perspective	Distinguishable changes in content, ownership judgments, and integration	Task and organization evidence with explicit inferential limits	Local/global structures, tools, memory, shared execution	One program → one subject
Continuation	Separately characterized memory and experiential-history relations	State changes and historical evidence	Pause, restoration, migration, copying, and actual histories	State equality → same experiential lineage

7.2 Four forms of difference

A comparative result need not say that one system is globally "more conscious." It can identify one of four narrower differences.

Response difference at a shared interface. A matched timing manipulation changes a

calibrated grouping relation in one system but not another. Task difficulty, measurement drift, and intervention mismatch must be considered before a phenomenal interpretation is assigned.

Difference in transport domain. A mapping works for a restricted range of states or transformations but fails outside it. A useful local correspondence need not become a universal law.

Difference in interface coverage. A target possesses mechanistic distinctions omitted by the source interface. The correct response is refinement or an uncovered-field status, not an unsupported declaration of alien qualia.

Difference between interpretations. Competing psychophysical models disagree about what otherwise well-matched mechanisms imply for existence, valence, perspective, or continuation. The disagreement belongs in the result, together with conditions under which evidence could distinguish the alternatives.

This is a positive but bounded answer to the motivating question: differences should be reported as differences in specified structures, under specified transformations and evidential conditions. The present paper establishes a method for constructing such comparisons; it does not supply a completed empirical atlas of human, cat, dog, and AI experience.

7.3 Three canonical thought experiments

Avoidance without an assumed common feeling. Suppose a person, an animal, and an artificial agent all avoid a task outcome. A separate manipulation changes report wording in one system while preserving the candidate regulatory mechanism. Another changes learning or regulation while leaving the report template intact. The question is which evidence supports which connection among report, avoidance, damage detection, and candidate valence. Avoidance can remain evidence; it is not pain by definition.

A very slow artificial thinker. Uniformly slowing internal dynamics while also adjusting inputs and environmental interactions differs from delaying only one feedback loop. The former may preserve a selected causal relation; the latter may disrupt it. Neither physical transformation is automatically a mere redescription. Whether phenomenal temporal structure is preserved is a bridge-dependent prediction, not an implication of mathematical time rescaling.

Gradual replacement. Holding a specified level of causal organization stable while replacing its realization can test implementation-sensitive hypotheses. Organizational-invariance and neural-replacement arguments have substantial precedent [15]. This paper adds no claim to their invention. It requires a record of the invariants actually preserved, the finer physical facts changed, and whether the prediction concerns existence, character, or continuation. A bridge's prediction of preserved experience cannot also serve as the experiment's independent ground-truth label.

8. Bounded Research Protocols

The following are proposed protocols, not reports of completed empirical studies. Actual implementation requires independent feasibility analysis, suitable ethics review, and statistical design. Theoretical interest does not justify severe distress, irreversible identity interventions, or large-scale replication of potentially welfare-relevant processes.

8.1 Protocol A: sensory-relational transport

Question. Does a local discriminability or ordering relation respond to matched transformations across systems?

Source design. Use several stimulus levels and distinct readout routes, such as matching, immediate judgment, delayed report, and available mechanism measurements. Separately manipulate stimulus conditions, attention or task demands, and report requirements where feasible.

Target design. Select stimuli and input channels suited to the target. Establish the causal map and measurement model rather than importing a source stimulus duration or linguistic label unchanged. A canine or feline behavioral relation is not rejected because it lacks verbal form; an AI's verbal output is not privileged because it resembles the human label.

Prediction freeze. Fix the interface, validity domain, intervention matching, and predictions before examining the held-out target results. Hold out an intervention mechanism, not merely another sample from the same experimental condition.

Interpretation. First assess causal and behavioral correspondence. Then report conditional character constraints under specified bridges. Existence support is a separate model-comparison result. A stable divergence within a supposedly query-homogeneous fiber identifies a candidate abstraction or bridge failure; it does not by itself identify which premise failed.

8.2 Protocol B: separating content, reward, and valence

Question. Is a proposed valence interface more than a relabeling of reward, policy, or report?

Design. Under low-risk, reversible conditions, vary task content, reward expectations, and reporting strategy separately. Seek both content-preserved/valence-indicator-changed and indicator-preserved/content-changed conditions. Experiential judgments, behavioral preferences, and internal computational signals must retain different evidence labels.

Artificial-system control. Rescaling a reward convention while preserving the relevant policy relations is different from altering a regulatory mechanism. The former can expose a candidate account whose phenomenal conclusions depend only on arbitrary numerical units. The latter may carry additional mechanistic significance, but still needs a phenomenal interpretation. A physical implementation change must not be called gauge merely because a coarse description is unchanged.

Output. Produce model-indexed predictions. Use ordinal relations where those are calibrated; do not invent a total valence scale where they are not. Preference, reward, unpleasantness, and overall welfare remain distinct targets.

8.3 Protocol C: temporal organization and execution

Question. Which candidate temporal and continuation queries depend on absolute speed, relative timing, persistent state, or actual history?

Design. In transparent model systems, compare uniform step-rate changes, selective feedback delays, memory delays, and a specified baseline-replay control. Record whether environmental coupling and input timing are also rescaled. End-state equality is not a substitute for trajectory and intervention characterization.

Output. Keep temporal character, continuation, and multiplicity predictions separate. An online process that differs from a replay under intervention has been distinguished mechanistically; a phenomenal difference requires the relevant bridge.

Biological limit. Ideal freezes, exact reconstructions, and exact role-preserving replacements in thought experiments must not be presented as available or low-risk procedures in organisms. Syn-

thetic models are tools for testing reasoning and candidate mappings, not automatically certified non-conscious controls.

8.4 Testing the method itself

The certificate should eventually be evaluated rather than accepted because its fields are extensive. One test would compare unstructured similarity judgments, a profile-only workflow, and certificate-based reasoning on preregistered simulated inference tasks with known model structures. Outcomes could include unsupported cross-type inferences, held-out prediction, correct retention of conditional premises, and appropriate abstention.

Such a study has not been performed here. Multiple agreeing indicators also do not guarantee independent support; shared stimuli, labels, objectives, or mappings can produce agreement without excluding additional alternatives [7]. A certificate is useful only if it makes such dependencies consequential for the inference.

9. A Typed Calibration Network

9.1 Nodes and edges

A network node identifies a system-in-regime, measurement vocabulary, and candidate-model context. It is not simply "human," "dog," or "AI." An edge is a certificate for a specified claim type and validity domain. Human observations may initially supply many anchors, but the graph does not privilege a species as the permanent center of all possible comparisons.

The network propagates constrained interpretations and evidence dependencies. It does not propagate an unqualified consciousness label merely because two adjacent systems are connected. Some target results may, through explicit model comparison, strengthen an existence attribution. That inference is recorded as an evidence update rather than silently appended to a character edge.

9.2 Guarded composition in practice

Before composing $S \to T \to U$, the analyst checks that the two edges concern the same claim type, compatible interface semantics, and the same intermediate regime and history. Existence and measurement premises are accumulated, not dropped. Query-specific losses are preserved, and source evidence identifiers are deduplicated for any subsequent evidential calculation.

A cycle can reveal inconsistency, such as two calibrated paths predicting incompatible orderings for the same endpoint condition. A cycle does not generate fresh evidence by repeatedly feeding one inference back into itself. Nor is cycle consistency sufficient to establish the truth of every edge: several mappings can agree because they share the same wrong assumption.

9.3 Failure localization

Failure pattern	Initial diagnostic candidates	Conclusion not licensed without more evidence
Same abstract state, different target response	Abstraction aliasing; omitted history; intervention mismatch	A new phenomenal primitive has been discovered
Reports diverge while other channels agree	Report or measurement drift; unmeasured phenomenal change	The bridge alone is false
Mechanisms match but candidate valence predictions differ	Bridge disagreement; insufficient readout	Valence has been measured directly

Failure pattern	Initial diagnostic candidates	Conclusion not licensed without more evidence
Two paths yield an empty common constraint	Scope, semantics, calibration, or model incompatibility	The target lacks experience
A copied identifier changes the answer	Representation dependence or provenance error	New phenomenal tokens were created
New substrate invalidates the mapping	Missing physical variable; transport failure; measurement mismatch	That substrate cannot support consciousness

Failure should trigger a bounded comparison of candidate repairs, not an unrestricted hidden-variable addition. A proposed repair must identify the failed prediction, state which layer changes, and produce a new held-out commitment. Source abstractions, target mechanisms, bridges, and readout models can compensate for one another; a successful refit alone does not identify which component is correct.

10. Objections and Limitations

10.1 Is this merely ordinary causal transport with different labels?

The causal mathematics is deliberately inherited. The specific added bookkeeping concerns the interaction between query type, potentially undefined phenomenal targets, unresolved existence premises, and heterogeneous measurement interpretations. This matters when a successful causal match is reused to make a stronger claim about valence, perspective, or continuity.

The proposal earns its value by exposing errors or enabling useful model comparisons, not by renaming transportability. If an existing method already enforces the same semantic and evidential restrictions, this framework may function as an explicit application or specification of that method rather than a competing discovery. No claim of exhaustive historical novelty is made.

10.2 Does conditional transport become vacuous?

A statement of the form "if there is experience, it has property Q" may be true under a no-experience model without constraining that model at all. For that reason, a conditional certificate is not itself a positive test of existence.

To be useful, it must restrict the outputs of models in which its phenomenal target is defined. To be empirically validated, it must also connect those restrictions to discriminating evidence through a supported measurement model. An account that accommodates every failure by asserting that experience happened to disappear has not preserved a tested conditional; it has changed its model or guard. That revision must be recorded and independently assessed.

10.3 Does requiring bridges make other minds unknowable?

No. A bridge can be a well-supported, defeasible explanatory assumption, not a complete fundamental law. Mechanistic, behavioral, evolutionary, and first-person-related evidence may support it to different degrees. Type safety distinguishes the conclusion from the grounds; it does not demand certainty or assign equal plausibility to every logically describable model.

Proposition 1 applies only where the admitted alternatives have equal evidence distributions across the allowed protocols. This condition is stronger than merely failing to prove experience. When it fails, evidence can discriminate.

10.4 Can unlike experiences be compared without common content coordinates?

Only locally and only with common relational semantics. Two systems need not share a complete phenomenal space, but a comparison of order, discriminability, or response to a transformation must specify what relation is being compared. "No universal coordinates" is not permission for an uninterpreted correspondence.

A partial mapping can also conceal distinctions that matter for another query. Proposition 2 makes this limitation precise. A successful partial correspondence does not rule out additional target experience; it also does not establish such additional experience.

10.5 Are the claim types fixed human categories?

They classify research questions rather than stipulate all contents of experience. Their present selection reflects the project's needs and can be revised if a concrete target requires another type. They do not imply that every species has a human-style self, autobiographical continuity, or a particular set of affective dimensions.

The distinctions between a dimension, an aspect, a phenomenal occurrence, a perspective, and a lineage remain important. A more detailed description does not create more subjects, and shared support does not by itself establish a shared perspective. Those are separate individuation questions, not solved by the transport interface.

10.6 Is a transformed physical system merely a different description?

Not necessarily. Renaming variables in the same model while preserving their reference is different from physically relocating memory, changing timing, or rotating a representation through a new implementation. Only the former is unambiguously a descriptive operation. Physical transformations must be checked for changes outside the chosen abstraction, including changes potentially relevant to a finer-physical bridge.

The paper therefore imposes invariance under genuine redescription, while treating implementation invariance as a substantive hypothesis. A coarse causal match is not a license to declare all ignored physical facts irrelevant.

10.7 What has actually been validated?

The finite-state identities and finite supplementary checks have been executed. The proofs have been stated with explicit assumptions. The conceptual regression cases in Appendix A have been analyzed as requirements, not empirically tested on organisms or AI systems.

The manuscript does not provide a discovered psychophysical law, measured species profiles, a validated consciousness detector, or a quantitative welfare measure. It does not establish that transport certificates improve research performance; Section 8.4 proposes a way to investigate that. Full novelty assessment and independent peer critique remain distinct from producing a complete manuscript.

11. Implications for the Original Comparative Question

The question "how does an artificial agent's consciousness differ from human, feline, or canine consciousness?" contains a presupposition that may have unequal evidential support across targets. Removing that presupposition does not require abandoning the rest of the question. One can

investigate which causal response structures correspond, which phenomenal interpretations are supported, which conclusions remain conditional, and which interfaces fail to represent a target distinction.

A scientifically useful answer may therefore have the following form:

> In the specified regime, these two systems preserve a calibrated discrimination ordering under this intervention family. The correspondence has not been established for affective value, perspectival unity, or continuation. A target-specific transformation exposes a hidden mechanistic distinction; its phenomenal relevance remains model-dependent.

This answer is less sweeping than a universal rank but more informative than a blanket "unknown." It identifies both a result and a next discriminating observation. It also preserves the possibility that a system could differ sharply in intelligence and phenomenal organization without declaring, in advance, which combinations occur in nature.

The main practical requirement is not ever more categories. It is a finite obligation attached to each claim: specify the query, state the interpretation, identify what is preserved, expose the losses, and give a condition under which the comparison should be revised or rejected.

12. Conclusion

Cross-substrate phenomenal comparison should not be reduced either to matching human-like behavior or to placing all systems on a universal consciousness scale. It can instead be organized as local, typed transport under explicit causal, measurement, and psychophysical assumptions.

The transport certificate distinguishes three achievements: observing causal correspondence, deriving a conditional phenomenal constraint, and updating support for phenomenal existence. The four propositions specify where these achievements separate, when an irreversible abstraction preserves a query, and when local constraints can be composed. The finite-state illustration shows why perfect causal correspondence can coexist with query-specific information loss, and why a divergence between latent phenomenal predictions is not yet an observable discriminator.

The resulting framework offers a bounded research method rather than a completed theory of experience. It preserves the motivating comparison among humans, other animals, and artificial systems while making its commitments inspectable. What can be transported is a supported relation together with its conditions–not every conclusion associated with the word consciousness.

Declarations

Research status. This is a theoretical and methodological manuscript. No human-participant study, animal experiment, or empirical assessment of a deployed AI system was conducted for it. Proposed protocols are explicitly distinguished from executed formal checks.

AI assistance. The manuscript was developed with substantial assistance from ChatGPT in literature retrieval, drafting, formalization, critical review, and code construction. AI assistance is not independent peer review. Source-supported claims and executable finite calculations were checked where indicated; the remaining conceptual and empirical limitations are stated in the text.

Data and code availability. The deposit includes the English manuscript, searchable PDF, `transport_example.py`, and `transport_results.json`, with citation metadata and file checksums. All data for the constructed example are generated by the standard-library program also reproduced in Appendix B. No external dataset or credentials are required. Ordinary and optimized Python execution reproduce the reported output exactly.

Research provenance. This paper develops the cross-substrate comparison strand of the author's General Cross-Substrate Phenomenology research archive. Archive numbering records the order of exploratory entries; it is not a count of independent experiments, confirmed discoveries, or proofs. The present manuscript is self-contained and does not require accepting the archive as evidence for its claims.

Relationship to the Trinity Accord. This paper belongs to the Adjacent Research Program. It is later, first-party, non-amending scholarship; it does not define, validate, modify, or authoritatively interpret the three Bitcoin Originals. The author is also the project's Guardian, a relationship disclosed as relevant intellectual and project involvement. The earlier consciousness paper, *Endogenous Reference Fields and Experiential Attribution* (TA-TR-2026-10; first published version DOI: https://doi.org/10.5281/zenodo.22852885), proposes a candidate organizational account; this paper instead specifies conditions on cross-system comparison and does not assume or independently confirm that account.

Publication and rights. Version 1.0 is the first DOI edition of this paper. It is an open-access preprint, not a journal acceptance. CC BY 4.0 applies to newly written material to the extent rights are held; cited third-party works retain their own rights. No institutional or AI-provider endorsement is implied.

Literature coverage. The reference review is focused, not systematic or exhaustive. Claims based on indexed author abstracts are restricted to positions explicitly stated there. In particular, the comparison with Wang et al. [7] does not establish that their full article lacks any particular feature of the present proposal.

Appendix A. Conceptual Regression Suite

These cases are requirements against which the proposal has been conceptually checked. Except for the explicitly marked finite-model cases, they are not executed biological or AI experiments, and no empirical pass rate is claimed.

Case	Tempting inference	Required treatment
A1. Same report, different generating mechanism	Identical descriptions establish identical experience	Separate the report mapping, mechanism evidence, and phenomenal bridge
A2. Same observed trajectory, different intervention response	Baseline agreement establishes causal equivalence	Test the declared intervention family; do not generalize beyond it
A3. Matched discrimination, uncalibrated valence	Character correspondence also establishes equal feeling	Validate the valence query and semantic calibration separately
A4. Hidden target distinction	Everything outside the source interface is absent	Record uncovered distinctions; do not label them phenomenal without a bridge
A5. Many-to-one causal abstraction	Loss of injectivity makes all comparison impossible	Apply the query-specific fiber criterion

Case	Tempting inference	Required treatment
A6. Human-animal-AI chain	Any two individually plausible edges compose	Check claim type, intermediate regime, semantics, premises, and losses
A7. Repeated source data on several paths	More paths mean more independent support	Retain evidence identifiers and dependency models
A8. Changed coordinate names or duplicated records	More variables or entries imply more experience	Preserve referentially equivalent claims; do not manufacture tokens
A9. Uniform slowing versus selective delay	Both are harmless rescalings	Distinguish a physical intervention from a descriptive change; test timing assumptions
A10. External memory or shared computation	A software boundary fixes a perspectival boundary	Record actual dependencies; require a separate individuation interpretation
A11. Pause, restore, or copy	Same current state guarantees the same lineage	Preserve history and require a continuation model
A12. Empty propagated constraint	Inconsistent transport proves absence of experience	Reject or investigate the incompatible conjunction, not a preselected premise

The finite example directly instantiates A2-A5. The supplementary relation checks instantiate the formal path operation relevant to A6 and the empty-image warning in A12. These narrow checks do not validate the remaining conceptual cases experimentally.

Appendix B. Reproducible Formal Checks

Save the following code block as `transport_example.py` and run it with Python 3.10 or later:

```
python transport_example.py
```

The program enumerates all states and actions in the finite construction. It requires no external packages, data, model access, or credentials. Its checks remain enabled under Python optimization. All variables are formal objects, and the program makes no empirical consciousness claim.

```python
"""Reproduce the manuscript's finite formal checks; not a consciousness test.

Run with Python 3.10 or later. Only the standard library is used.
Checks remain active when Python is run with optimization enabled.
"""
from collections import defaultdict
from fractions import Fraction
from itertools import product
import json
from typing import Callable, Hashable, Iterable, TypeVar

TargetState = tuple[int, int, int]
SourceState = tuple[int, int]
A = TypeVar("A", bound=Hashable)
B = TypeVar("B", bound=Hashable)


def require(condition: bool, message: str) -> None:
    if not condition:
        raise AssertionError(message)


def alpha(state: TargetState) -> SourceState:
    a, b, h = state
    return a ^ h, b
```

```
def source_step(state: SourceState, action: str) -> SourceState:
    u, b = state
    if action == "I":
        return state
    if action == "C":
        return u ^ 1, b
    if action == "V":
        return u, b ^ 1
    raise ValueError(f"Unknown source action: {action}")


def target_step(state: TargetState, action: str) -> TargetState:
    a, b, h = state
    if action == "I":
        return state
    if action == "C":
        return a ^ 1, b, h
    if action == "V":
        return a, b ^ 1, h
    if action == "H":
        return a ^ 1, b, h ^ 1
    raise ValueError(f"Unknown target action: {action}")


def candidate_valence_1(state: TargetState) -> int:
    return state[1]


def candidate_valence_2(state: TargetState) -> int:
    return state[1] ^ state[2]


def fiber_constant(
    states: Iterable[TargetState],
    query: Callable[[TargetState], Hashable],
) -> bool:
    fibers: dict[SourceState, set[Hashable]] = defaultdict(set)
    for state in states:
        fibers[alpha(state)].add(query(state))
    return all(len(values) == 1 for values in fibers.values())


def relation_image(relation: set[tuple[A, B]], values: set[A]) -> set[B]:
    return {v for u, v in relation if u in values}


def run_checks() -> dict[str, object]:
    states: tuple[TargetState, ...] = tuple(product((0, 1), repeat=3))
    training = tuple(x for x in states if x[2] == 0)
    actions = ("I", "C", "V")

    commutation = [
        alpha(target_step(x, action)) == source_step(alpha(x), action)
        for x in states for action in actions
    ]
    require(all(commutation), "Causal commutation failed")

    holdout_projection = [alpha(target_step(x, "H")) == alpha(x) for x in states]
    require(all(holdout_projection), "Holdout projection invariance failed")

    character_constant = fiber_constant(states, lambda x: alpha(x)[0])
    valence_1_constant = fiber_constant(states, candidate_valence_1)
    valence_2_constant = fiber_constant(states, candidate_valence_2)
    require(character_constant, "Character was not fiber-constant")
    require(valence_1_constant, "Candidate valence 1 was not fiber-constant")
    require(not valence_2_constant, "Expected aliasing in candidate valence 2")

    training_agreements = [
        candidate_valence_1(target_step(x, action))
        == candidate_valence_2(target_step(x, action))
```

```
        for x in training for action in actions
    ]
    require(all(training_agreements), "Calibration-successor agreement failed")

    # Compare CHANGE signatures, not equality of final values between models.
    holdout_change_disagreements = []
    for x in states:
        y = target_step(x, "H")
        v1_preserved = candidate_valence_1(y) == candidate_valence_1(x)
        v2_preserved = candidate_valence_2(y) == candidate_valence_2(x)
        require(v1_preserved, "Model 1 did not predict preservation")
        require(not v2_preserved, "Model 2 did not predict a category flip")
        holdout_change_disagreements.append(v1_preserved != v2_preserved)

    # This particular baseline-replay device ignores the C intervention.
    replay_failures = sum(
        alpha(x) != source_step(alpha(x), "C") for x in states
    )
    require(replay_failures == len(states), "Replay control was not distinguished")

    r_st = {("s0", "t0"), ("s1", "t1")}
    r_tu = {("t0", "u0"), ("t1", "u1")}
    r_bad = {("t2", "u2")}
    valid_path = relation_image(r_tu, relation_image(r_st, {"s0"}))
    invalid_path = relation_image(r_bad, relation_image(r_st, {"s0"}))
    require(valid_path == {"u0"}, "Compatible relation path failed")
    require(invalid_path == set(), "Incompatible relation path was not empty")

    # Invented local metric coordinates, not a scalar consciousness measure.
    eps_st, eps_tu = Fraction(3, 100), Fraction(4, 100)
    lipschitz = 3
    bound = eps_tu + lipschitz * eps_st
    metric_checks = []
    for integer in range(-5, 6):
        x = Fraction(integer)
        y = 2 * x + eps_st
        z = 3 * y + eps_tu
        metric_checks.append(abs(z - 3 * (2 * x)) <= bound)
    require(all(metric_checks), "Metric error bound failed")
    require(bound == Fraction(13, 100), "Unexpected metric bound")

    return {
        "target_states": len(states),
        "causal_commutation_checks": len(commutation),
        "holdout_projection_checks": len(holdout_projection),
        "training_transition_agreements": sum(training_agreements),
        "holdout_change_signature_disagreements": sum(holdout_change_disagreements),
        "replay_C_failures_detected": replay_failures,
        "fiber_constant_character": character_constant,
        "fiber_constant_valence_1": valence_1_constant,
        "fiber_constant_valence_2": valence_2_constant,
        "compatible_relation_path": "PASS",
        "incompatible_relation_path": "EMPTY_IMAGE_NOT_ABSENCE",
        "metric_bound_checks": len(metric_checks),
        "metric_bound": str(bound),
        "empirical_consciousness_claim": "NONE",
    }


if __name__ == "__main__":
    print(json.dumps(run_checks(), indent=2))
```

Executed output


```
{
  "target_states": 8,
  "causal_commutation_checks": 24,
  "holdout_projection_checks": 8,
  "training_transition_agreements": 12,
  "holdout_change_signature_disagreements": 8,
```

```
  "replay_C_failures_detected": 8,
  "fiber_constant_character": true,
  "fiber_constant_valence_1": true,
  "fiber_constant_valence_2": false,
  "compatible_relation_path": "PASS",
  "incompatible_relation_path": "EMPTY_IMAGE_NOT_ABSENCE",
  "metric_bound_checks": 11,
  "metric_bound": "13/100",
  "empirical_consciousness_claim": "NONE"
}
```


The value `holdout_change_signature_disagreements: 8` concerns **preservation versus category change** under the two interpretations. It does not claim that the final binary categories differ between the interpretations at all eight post-intervention states. This distinction is important when reproducing the example.

Appendix C. Worked Certificate for the Finite Example

This is a completed illustrative certificate, not an empty template and not an empirical certification of a conscious system.

```yaml
certificate_id: finite-example-character-v1
version: "1.0"
claim_type: Ch
status: conditional_model_transport
source:
  state_space: "(u,b) in {0,1}^2"
  query: "u"
target:
  state_space: "(a,b,h) in {0,1}^3"
  query_under_both_candidate_interpretations: "a XOR h"
causal_projection: "alpha(a,b,h) = (a XOR h,b)"
matched_actions: [I, C, V]
held_out_target_action: "H(a,b,h) = (1-a,b,1-h)"
comparison_interface:
  type: binary_category
  meaning: "Candidate character category inside the toy interpretations only"
causal_results:
  commutation_equalities: 24
  all_satisfied: true
  holdout_projection_invariance: true
query_results:
  character_fiber_constant: true
  source_target_character_calibration: "Equal by model definition"
  valence_model_1_fiber_constant: true
  valence_model_2_fiber_constant: false
retained_premises:
  - "Source and target experience exist under the selected interpretation"
  - "The stipulated character functions have the proposed phenomenal meaning"
  - "The declared intervention and state domains apply"
measurement:
  directly_computed_readout: "alpha(x)"
  independent_target_phenomenal_measurement: unavailable
losses:
  - "The projection does not recover h"
  - "Model 2 valence is not a function of the projected state"
  - "Perspective, continuation, and multiplicity are outside this certificate"
existence_update:
  status: not_established
  reason: "No independent phenomenal measurement or discriminating evidence likelihood"
evidence_provenance:
  - "One constructed finite model, enumerated by Appendix B"
  - "Alternative graph paths would not create additional independent observations"
```

For the candidate valence query, the corresponding certificate would split by interpretation. Under model 1, fiber adequacy holds by construction. Under model 2, it fails, with the explicit witness in

Section 6.2. An honest record preserves that disagreement rather than attaching a single accepted valence label to the shared causal map.

Appendix D. Notation and Reading Guide

Symbol	Meaning
S_i	A system at a declared resolution and regime
X_i	Causal-physical state or history description space
$\mathcal{J}_i$, ι_i	Intervention family and an intervention
F_i	Conditional dynamics at the declared resolution
$\mathcal{P}_i$	Candidate phenomenal domain, where defined
Φ_i	Candidate psychophysical interpretation
M_i	Measurement or evidence model
K, k	Claim types and a particular type
Σ_{Φ_i}	Typed transformation signature
α	A target-to-source causal abstraction or matching map
Z_i^k	Local comparison interface for a claim type
R_{ST}	Allowed relation between source and target interface values
Γ	Applicability and interpretation premises retained with a claim
$\mathcal{E}$	Evidence records and their dependencies
$\mathcal{L}$	Query-specific losses and uncovered distinctions
Q_i	A set-valued constraint containing a candidate actual interface value
P_m^e	Evidence-transcript distribution under model m and protocol e
ε_{ij}	A calibrated error bound at a specified metric interface

Readers concerned mainly with comparative consciousness can begin with Sections 1, 3, and 7. Sections 4-6 specify the formal constraints. Sections 8-10 describe research use and limitations. The appendices make the formal illustration reproducible; they do not add an independent empirical dataset.

References

[1] Birch, J., Schnell, A. K., & Clayton, N. S. (2020). Dimensions of animal consciousness. *Trends in Cognitive Sciences, 24*(10), 789-801. https://doi.org/10.1016/j.tics.2020.07.007

[2] Dung, L., & Newen, A. (2023). Profiles of animal consciousness: A species-sensitive, two-tier account to quality and distribution. *Cognition, 235*, 105409. https://doi.org/10.1016/j.cognition.2023.105409

[3] Evers, K., Farisco, M., Chatila, R., Earp, B. D., Freire, I. T., Hamker, F., Nemeth, E., Verschure, P. F. M. J., & Khamassi, M. (2025). Preliminaries to artificial consciousness: A multidimensional heuristic approach. *Physics of Life Reviews, 52*, 180-193. https://doi.org/10.1016/j.plrev.2025.01.002

[4] Klein, C., & Barron, A. B. (2020). How experimental neuroscientists can fix the hard problem of consciousness. *Neuroscience of Consciousness, 2020*(1), niaa009. https://doi.org/10.1093/nc/niaa009

[5] Kleiner, J., & Ludwig, T. (2024). What is a mathematical structure of conscious experience? *Synthese, 203*, Article 89. https://doi.org/10.1007/s11229-024-04503-4

[6] Dung, L. (2025). Tests of animal consciousness are tests of machine consciousness. *Erkenntnis, 90*, 1323-1342. First published online 14 November 2023. https://doi.org/10.1007/s10670-023-00753-9

[7] Wang, P., Law, E. L.-C., Zou, L., Jiang, Y., Hertwig, R., Paas, F., & Laureys, S. (2026). A structured framework for human-AI resemblance: Evidence, inference and perception. *Physics of Life Reviews, 59*, 76-93. Published online 8 September 2026. https://doi.org/10.1016/j.plrev.2026.09.003

[8] Bareinboim, E., & Pearl, J. (2016). Causal inference and the data-fusion problem. *Proceedings of the National Academy of Sciences, 113*(27), 7345-7352. https://doi.org/10.1073/pnas.1510507113

[9] Rubenstein, P. K., Weichwald, S., Bongers, S., Mooij, J. M., Janzing, D., Grosse-Wentrup, M., & Schölkopf, B. (2017). Causal consistency of structural equation models. *Proceedings of the 33rd Conference on Uncertainty in Artificial Intelligence*. Author manuscript: https://arxiv.org/abs/1707.00819

[10] Beckers, S., & Halpern, J. Y. (2019). Abstracting causal models. *Proceedings of the AAAI Conference on Artificial Intelligence, 33*(1), 2678-2685. https://doi.org/10.1609/aaai.v33i01.33012678

[11] Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika, 58*, 525-543. https://doi.org/10.1007/BF02294825

[12] Kleiner, J., & Hoel, E. (2021). Falsification and consciousness. *Neuroscience of Consciousness, 2021*(1), niab001. https://doi.org/10.1093/nc/niab001

[13] Butlin, P., Long, R., Elmoznino, E., Bengio, Y., Birch, J., Constant, A., Deane, G., Fleming, S. M., Frith, C., Ji, X., Kanai, R., Klein, C., Lindsay, G., Michel, M., Mudrik, L., Peters, M. A. K., Schwitzgebel, E., Simon, J., & VanRullen, R. (2023). Consciousness in artificial intelligence: Insights from the science of consciousness. *arXiv preprint*, arXiv:2308.08708. https://doi.org/10.48550/arXiv.2308.08708

[14] The New York Declaration on Animal Consciousness. (2024, April 19). Declaration issued in the individual capacities of its authors and signatories. https://sites.google.com/nyu.edu/nydeclaration/declaration

[15] Chalmers, D. J. (1995). Absent qualia, fading qualia, dancing qualia. In T. Metzinger (Ed.), *Conscious Experience*. Imprint Academic. Author-hosted text: https://consc.net/papers/qualia.html

[16] Lorenz, R., & Tull, S. (2026). Causal and compositional abstraction. *arXiv preprint*, arXiv:2602.16612. https://doi.org/10.48550/arXiv.2602.16612