Article

Not peer-reviewed version

Steins Theory: A New Axiomatic System for Identity

Jiaqi Guo *

Posted Date: 23 December 2025

doi: 10.20944/preprints202506.2384.v7

Keywords: ethics; identity; category mistake; n = n; quantum identical particles; spacetime leap

 doi:10.20944/preprints202506.2384.v7

Article

Steins Theory: A New Axiomatic System for Identity

Jiaqi Guo

Independent Researcher, China; 1028165464@qq.com

Abstract

In the philosophy of language, Frege's (1892) distinction between sense and reference provided a foundational framework for identity statements, while Putnam's (1975) "Twin Earth" thought experiment, with its astonishing insight, pushed externalism to its extreme, successfully challenging internalist models of meaning and setting the basic agenda for debates on reference determination for decades to come. However, despite the inspirational nature of these groundbreaking works, a puzzling phenomenon emerges: the debates they sparked—such as discussions around core cases like the Ship of Theseus and identical particles—appear to have fallen into a kind of impasse. This paper argues that this impasse may not stem from the depth of the problems themselves, but precisely from an unexamined deep assumption shared by these otherwise highly persuasive theories: namely, the belief that there exists some single, decisive level (whether microscopic physical structure or historical causation) that can once and for all resolve the identity question. This paper proposes that, rather than continuing to seek a superior single answer under this assumption, a more productive path may be to reflect on the assumption itself. To this end, we develop a framework of hierarchical relativity. Interestingly, this framework shows that the aforementioned seemingly opposing outstanding theories can actually be understood as special cases of this framework at different levels; the difficulties they encounter become inevitable precisely when they attempt to make assertions across levels. Thus, this framework is not intended to negate prior work, but to clarify its valid scope of application, providing a new path to crack a series of philosophical puzzles born from category mistakes.

Keywords: ethics; identity; category mistake; $n \equiv n$; $n \neq m$; quantum identical particles; logical fallacy; spacetime leap

1. Introduction

The problem of identity, that is, "what makes a thing what it is," lies at the core of metaphysics and logic. From Leibniz's (Gottfried Wilhelm Leibniz, 1714) profoundly insightful Principle of the Identity of Indiscernibles (PII)—which, with its logical simplicity and power, set the lofty ideal for the individuation of entities—to Kripke's (Saul Kripke, 1980) groundbreaking theory of necessary identity based on rigid designators and origins, which, with its deep grasp of modal intuitions, provided a seemingly solid foundation for the stability of reference. These outstanding efforts of generations of philosophers have collectively constructed the grand intellectual tradition for our understanding of individual persistence and recognition.

However, a puzzling phenomenon is that these theoretical frameworks, each highly persuasive in their respective domains, often exhibit a regrettable, systematic limitation when confronting complex boundary cases posed by reality and thought experiments. Leibniz's strong PII encounters fundamental difficulties perhaps unforeseen in its original conception when facing quantum identical particles (French & Redhead, 1988). The intrinsic rigor of the principle instead plunges it into a near-paradoxical impasse when dealing with multiple entities (such as two electrons) that are completely indistinguishable in all intrinsic properties yet numerically distinct. Similarly, Kripke's elegant theory, aimed at anchoring reference, becomes quite entangled in explaining diachronic changes in intrinsic properties (such as the Ship of Theseus) (Chandler, 1975), not to mention its difficulty in

 doi:10.20944/preprints202506.2384.v7

accommodating the identical particle phenomena presented by quantum mechanics that challenge the classical view of individuality.

Thus, we face a peculiar intellectual impasse: whether Leibniz's strong program pursuing identity of properties or Kripke's historical path focusing on the necessity of origins, they seem to successfully illuminate one flank of the edifice of identity, yet unfortunately leave the other side in deeper shadows. Their shared, perhaps unarticulated ambition—to find a single, absolute criterion for identity—itself, may precisely constitute an a priori obstacle, preventing us from truly understanding the multidimensional nature of the problem?

This paper argues that the key to breaking this impasse is not to make an either-or choice among existing paths or to engage in another round of patching. Instead, it requires us to step back and conduct a meta-level reflection on the problem itself. This paper aims to propose a hierarchical relativity framework for identity analysis. The ambition of this framework is not to completely negate predecessors—in fact, Leibniz's attention to properties and Kripke's obsession with history will be repositioned and gain their limited legitimacy in the new framework—but to show that the root of the aforementioned dilemmas lies in a category mistake: that is, erroneously attempting to answer a question at one level with an answer from another level.

This paper aims to transform classical controversies in ontology and semantics into a clear, operable conceptual choice problem. Thus, it does not solve these puzzles but dissolves them, providing a new way out for a series of philosophical anxieties born from category mistakes.

Note: Unless a proposition explicitly specifies premises (e.g., considering from a certain angle), this paper defaults to premises that are effective in reality (physics). This default is because the paper considers that if a proposition does not pursue real-world effectiveness, then a default shift in perspective naturally makes the truth or falsity of the proposition's conclusion unimportant.

2. Analysis

2.1. Axiom 1 (Self-Identity)

$\forall n \in I, n \equiv n$, (n) is necessarily identical to itself, which is the foundation of logical reference.

2.2. Axiom 2 (Mutual Distinction)

$\forall n \forall m (n \neq m \rightarrow \exists F (F(n) \wedge \neg F(m)))$

Core Corollary:

Uniqueness Theorem: Two (n) with identical content must be the same (Citation 1)

Note: See Section 3.1 for examples.

2.3. Category Mistakes and Level Confusion: The Ship of Theseus as an Example

In addressing the enduring puzzle of the Ship of Theseus, a highly influential and intuitively attractive solution has been proposed, represented by David Wiggins (1980). This solution asserts that the identity of an object is not guaranteed by its properties at any single moment, but must be borne by its continuity in spacetime and an uninterrupted historical causal path. The great advantage of this approach is that it successfully captures our deep intuition about "objects persisting through time"—that things are not instantaneous existences but have their life histories (careers) or biographies.

Another solution, perdurantism (four-dimensionalism), provides a radically different and metaphysically elegant picture (see Heller, 1984; Sider, 2001). With its thorough clarity, this theory ingeniously avoids many traps of diachronic identity. Perdurantism asserts that the Ship of Theseus is not a three-dimensional entity that "fully exists" through time, but a "spacetime worm" extended in four-dimensional spacetime. Each time slice of the ship is regarded as a temporal part of this four-dimensional object. Thus, so-called "change" merely means that this four-dimensional whole has different properties (such as different planks) at different temporal parts. In this framework, the

 doi:10.20944/preprints202506.2384.v7

original Ship of Theseus (as a four-dimensional entity), the replaced ship, and the ship reassembled from the old planks are three different four-dimensional objects. They may have completely identical three-dimensional cross-sections at some time segment (thus indistinguishable at that moment), but as wholes, they are naturally different. The excellence of this solution lies in transforming the problem of persistence from the troubling "identity" to the relatively clear "part-whole" relation.

Closely related to perdurantism is stage theory (see Sider, 1996; Hawley, 2001), which, while retaining many advantages of perdurantism, attempts to better accommodate our everyday language intuition that "objects are three-dimensional." Stage theorists argue that what we usually call "the Ship of Theseus" refers not to the entire four-dimensional worm, but to one of its stages or time slices at a specific time point. When we say at time t that "the ship is the same," we are actually saying that there exists a (primitive) counterpart relation between the stage at t and the stage at an earlier time t_1, maintained by some similarity and causal continuity. Stage theory, with its conceptual economy, avoids presupposing identity relations across time, thus exhibiting strong theoretical appeal.

However, although the above theories are each ingeniously crafted in their internal logic and often self-consistent, this paper will argue that they face a common, profound dilemma at the normative level. Whether appealing to historical paths, four-dimensional wholes, or counterpart relations, these theories all attempt to provide a single, absolute criterion for identity judgment. To achieve this goal, they must construct "time" or "spacetime position" itself as a constitutive element of the object's identity. This means that within their theoretical frameworks, answering the question "Is the ship at t_1 the same as at t_2?" logically necessarily depends on checking spacetime coordinates or cross-time associations.

This paper argues that it is precisely this key theoretical move that inadvertently leads to a "category mistake." Let us formally reconstruct the solutions of each theory: they actually adopt a domain defined as (ship's physical properties, historical causal path/four-dimensional whole/counterpart relation).

Let us apply n ≡ n to the Ship of Theseus. First, we must clarify the question: when we ask "Is the replaced ship still the original ship?", what is the default comparison category? A reasonable interpretation is that we care about its identity as an "objectively verifiable ship," that is, the category (physical ship), with relevant properties {e.g., all particles of the ship and their arrangement shape}.

Now, let us examine the historical path theory. This theory, in answering the above question, actually introduces a new category (physical ship, history), with relevant properties {all particles of the ship and their arrangement shape, historical causal path}. In the (physical ship, history) category, since the historical path has changed, the replaced ship is naturally different from the original ship.

This paper argues that the controversy here stems from confusion of categories. The questioner implicitly asks within the (physical ship) category, while the historical path theory answers within the (physical ship, history) category. These two answers—"yes" {in (physical ship)} and "no" {in (physical ship, history)}—do not contradict because they answer two different questions. Imposing the answer from (physical ship, history) onto the question from (physical ship) constitutes a category mistake. The true value of historical path theory lies in revealing "history" as an important category, but it erroneously treats it as the sole decisive category.

We can formally reconstruct this as follows:

1. Initial question: Determine whether the entity "ship" at time points t_1 and t_2 is the same.

2. Correct (n) domain (based on the initial question): Should include only properties related to the "ship" substance, i.e., physical ship: (all particles of the ship and their arrangement shape)

3. Category mistake in historical/four-dimensional/stage theories:

· Wiggins actually adopts history + physical ship (all particles of the ship and their arrangement shape, historical causal path).

· Perdurantism actually adopts 4D + physical ship (all particles of the ship and their arrangement shape, spacetime coordinates).

· Stage theory actually adopts stage + physical ship (all particles of the ship and their arrangement shape, counterpart relation).

 doi:10.20944/preprints202506.2384.v7

The process leading to this can be formally expressed as:

· They attempt to answer the identity question based on (physical ship): "Is Ship at t1 ≡ Ship at t2?"

· However, the judgment domain they actually use is (physical ship + history, 4D, or stage).

· Since the materials in the replacement process are identical to the original materials, we have (physical ship_t1) = (physical ship_t2).

· But because the historical path, spacetime coordinates, or counterpart relation has changed, (physical ship_t1) ≠ (physical ship + history/spacetime/stage_t2).

· Thus, they conclude Ship_t1 ≠ Ship_t2 to answer the physical ship question.

Thus, we see an interesting situation: each theory effectively answers a question, but possibly not the one initially posed. They precisely answer "Is there a continuous four-dimensional worm connecting the ship at t_1 and t_2?" or "Is the stage at t_2 the counterpart of the stage at t_1?", but treat this answer as the ultimate adjudication of the question "Is the ship structurally the same?" This is like a judge, asked "Does the defendant's behavior comply with Criminal Law Article X?", giving a verdict after consulting the civil law code. The conclusion may be coherent within its own system, but it has quietly switched the venue of the debate, which is a category mistake.

Therefore, the true contribution of historical path theory may lie in its excellence in revealing how various explanatory properties such as "history" and "spacetime whole" influence our identity judgments. But its limitation is that it attempts to elevate such explanatory properties to metaphysical necessities, thus having to expand the criteria for identity questions to maintain the integrity of its theory. The value of this theory is that it does not need to make this difficult expansion, but by clarifying the levels of the question, allows different domains/properties to give effective and non-conflicting answers to different questions. It does not solve these puzzles but dissolves them.

2.4. Conservation

2.4.1. The Indistinguishability of Identical Particles in Quantum Mechanics Poses the Most Severe Challenge to Leibniz's PII (Citations 1,2), Yet Provides a Natural, Physically Evidenced Model for This Theory

Current philosophical discussions on identical particles, facing the challenge of quantum identical particles to PII, mainly fall into two categories of mainstream solutions: revisionism and revolutionism. The former attempts to salvage some form of individuality principle, while the latter abandons individuality itself.

Saunders's solution undoubtedly represents one of the most ingenious and technically rigorous attempts in the revisionist path. By ingeniously defining "weak discernibility," he successfully liberates the discussion from the dead end of intrinsic properties, providing an insightful perspective for finding the cornerstone of individuation in relations. The complexity of this approach and the extensive discussions it has sparked in itself prove its profound philosophical value. However, it is precisely this technical complexity that exposes a potential cost underlying the solution: its definition of "purely extensional relational properties," though striving for precision, inevitably introduces considerable terminological ambiguity, to the extent that its defenders must also carefully handle accusations of circular argumentation (Muller & Saunders, 2008). More centrally, the entire theoretical edifice of this solution is built on an unsettling presupposition: that the "individuality" of identical particles must, and can only, be "saved" by finding some (even relational) individuating property. This makes its theoretical efforts—no matter how ingenious—essentially an ad-hoc repair to salvage a premise. When applied to states with indefinite particle number in quantum field theory, this strategy of continuously introducing new relational properties to salvage individuality becomes increasingly ad-hoc: it no longer resembles an elegant deduction of the theory itself, but more like an increasingly costly price paid to maintain the theoretical premise (that individuality must exist).

Facing the dilemmas of revisionism, another revolutionary solution chooses a more thorough path. Scholars represented by Décio Krause (2011) propose a highly subversive argument: quantum particles may not be "individuals" in the traditional metaphysical sense at all. Therefore, laws based

 doi:10.20944/preprints202506.2384.v7

on individual identity are categorically wrong from the start. They should be understood as "non-individuals" and described using highly specialized mathematical tools such as quasi-set theory.

Krause's solution is striking for its conceptual thoroughness and consistency; it unreservedly embraces the most counterintuitive features of quantum mechanics, decisively breaking with our entire classical framework of objects and spacetime positioning. This resolute posture is undoubtedly clean and efficient in theory. However, the corresponding cost of this efficiency is that the concept of "non-individual" itself imposes considerable explanatory burden in metaphysics, requiring us to abandon an entire set of deeply rooted intuitive understandings of what "one thing" means.

2.4.2. This Paper's Solution: A Hierarchical Relativity Framework

The above two solutions share a deep misconception: they both attempt to find answers to a wrongly posed question. The problem is not "What is the correct individuating property?", but "At what category are we inquiring about the identity question?"

This paper provides a meta-framework for this. We define an observable particle state as: P = (o, q), where o is the set of intrinsic properties (mass, charge, spin, etc.), and q is the set of spacetime coordinates.

· When we inquire at the level of (particle) = o, i.e., comparing only intrinsic properties, all identical electrons are (electron) = (mass m_e, charge -e, spin 1/2...). According to Axiom 1, at this level, they are indeed the same electron e. This explains the root of their indistinguishability.

· When we inquire at the level of (particle state) = (o, q), since q (such as position) is necessarily different, $(P_1) \neq (P_2)$, so they are different particle states. This explains why we observe multiple scattering events in experiments.

Thus, the confusion brought by quantum identical particles stems from erroneously invading the difference at the q (spacetime coordinates) level into the identity judgment at the o (intrinsic properties) level. Steins Theory resolves the contradiction by clearly distinguishing these two levels: they are both "one" (as a logical concept) and "many" (as manifestations in specific spacetime). Particle annihilation and creation merely represent the decoupling and re-coupling of e with different coordinates q.

The advantage of this solution is that it absorbs the advantages of Krause's solution in acknowledging the specificity of quantum mechanics (by interpreting "non-individuality" as identity at the o level), while avoiding its radical metaphysical cost (we are still talking about "quanta," just in different categories); at the same time, it explains why Saunders's strategy of introducing relational properties seems feasible in some cases (because he erroneously took q-level properties as the basis for individuation at the o level), yet fundamentally misguided.

2.4.3. Formal Derivation Proof of Conservation:

Let the basic particle state be expressed as: Particle P = (o, q) where:

· o is the set of intrinsic properties (such as mass m, charge q, spin s)

· q is the set of spacetime coordinates (such as position x, time t).

Formalization:

1. When the domain of p is P = (o, q_1), a certain coordinated electron

2. Coordinate decoupling (destruction): (o, q_1) → (o), (q_1) ⇒ The particle degenerates to a pure eigenstate (o), unmeasurable due to lack of observable basis (q = ∅). ∀ particle states (o, q_1) and (o, q_2), it can be found that: (o) ≡ (o) indicates:

· When the eigenproperties of two particles are indistinguishable (o ≡ o), regardless of how their spacetime coordinates $q_1 \neq q_2$ differ, their particle is the same electron **e** = (o) projected in different spacetimes

Physical interpretation:

· Particle annihilation ⇨ Set decoupling rather than extinction ⇒ **e** = (o) enters a free state

 doi:10.20944/preprints202506.2384.v7

· Particle creation ⇨ The same **e** binds new coordinate q_2 ⇒ Observed as reappearance, example: Electron *e* disappears at position q_1 and appears at q_2, actually the coordinate migration of electron **e** = (*q*=-1e, *m*_e, *s*=1/2...): (**e**, q_1) → (**e**) → (**e**, q_2), its electron identity guaranteed by n ≡ n.

Direct corollary: Conservation theorem: What logic allows exists, will not annihilate nor update

2.5. Symmetry

Max Black's (1952) symmetric universe thought experiment poses the most extreme challenge to Leibniz's strong PII. He imagines a universe with only two completely identical spheres. These two spheres are indistinguishable in all intrinsic properties (mass, composition, shape, etc.) and all relational properties (distance X miles apart, symmetric to each other). Black thereby argues that this is a genuine scenario of "two" things, thus refuting PII—there is no property to distinguish them, yet they are still numerically two distinct entities.

Traditional response strategies mainly fall into two types: one questions the metaphysical possibility of such a symmetric universe (e.g., requiring a basis for "numerical difference" itself, which usually loops back to some hidden property); the other, like Saunders (2003), argues that relational properties (such as "being X miles from a sphere") can themselves serve as weakened distinction bases. However, the former is criticized as ad-hoc, while the latter is difficult to work in Black's original setting because each sphere's relational properties ("being X miles from another sphere") remain completely identical.

This paper argues that Black's challenge and the dilemmas of traditional responses jointly root in an unexamined presupposition: that "numerical two" is a primitive, irreducible fact. This theory provides a brand-new analytical perspective. Under this paper's framework, we must first clarify the domain of (n).

· If (sphere) is defined as the set of all traditional properties (intrinsic + relational), i.e., (n) = {mass M, shape spherical, ..., distance X from a sphere}, then according to the axiom, since (n) ≡ (n), we inevitably conclude sphere ≡ sphere. This seems to directly derive the PII conclusion that Black attempted to refute.

· However, Black's intuition—"there are clearly two spheres here"—is not entirely baseless. This theory interprets it as a fixed pattern in thinking. The reason observers report "seeing two" is because their perspective itself is embedded in this symmetric spacetime coordinate system. This paper argues that a fundamental misconception shared by Black and his commentators is assuming that the reference of "sphere" and "sphere" necessarily corresponds to two entities with independent spacetime coordinates. This presupposition forces them into a dilemma between "abandoning PII" or "inventing new metaphysical concepts." The concept of "Coordinate Self-Reference" provides a third way out of this dilemma. Formalization: For the entire symmetric system S, define: (S) = {there exists a sphere with property set P, and the sphere is opposite itself}. This description looks complex, but simply put, (S) describes a single coordinate framework that allows "self-facing." Within this framework, a sphere being opposite itself is not a grammatical error, but an accurate description of a singularity coordinate topology. Visually, the "two" spheres presented are projections of this single, self-referential coordinate structure in Euclidean space perception (similar to an object and its mirror image, but here there is no mirror; it is the topological property of space itself).

· System S: (S) describes the state after a single sphere is bound to a special self-referential coordinate topological structure: (sphere, R_self-facing).

· Paradox dissolution: Black's error lies in erroneously inferring the existence of two spheres (sphere_1, sphere_2) from the system state (sphere, R_self-facing). He confused categories, using the description result of (S) to answer the question about (single sphere). In fact, there never existed a second sphere; what always existed was only one sphere, in a special coordinate topology that produces a "double image projection."

Thus, this framework does not deny our intuition of "seeing two spheres," but provides a brand-new, more precise ontological explanation for this intuition: it is the perception of a single sphere in a self-referential coordinate topology. This successfully resolves the apparent contradiction between

 doi:10.20944/preprints202506.2384.v7

PII and counting intuition, while avoiding the introduction of any ad-hoc individuating factors. Black's challenge not only fails to refute the law of identity but, through the level analysis of this framework, more profoundly reveals the dependence of "identity" judgments on the background framework.

3. Examples

3.1. Copy Paradox

· Controversy: Two documents with identical content stored on different devices, are they two pieces of information?

· Solution:

· If the goal is pure content identity → (n) = text semantics, then n ≡ n;

· If the goal is document location entity identity → (n) = (text semantics, location), then (content, Loc_A) ≠ (content, Loc_B).

· Conclusion: Copies are observable due to the set formed by the same information and different spacetime coordinates.

3.2. Gibbs Paradox

Category mistake:

· The goal should be particle type identity → (n) = (mass, spin,...)

· Classical statistics privately expands to (n) = (intrinsic properties, fictional labels)

Correction: (n) and (n, labels) ⇒ (n) ≡ (n) Entropy increase error stems from erroneous choice of (n) domain (introducing labels)

3.3. Black Hole Information Paradox

Category mistake: Binding the domain defined as internal properties (n) to spacetime coordinates (n) = (information structure, black hole coordinates)

Correct solution:

· Define goal: Internal property identity → (n) = quantum properties

· Black hole disassembles the set (quantum properties, coordinates), uncoordinated content leads to unobservability, but (quantum properties/coordinates) as a logical concept does not disappear

· If new spacetime satisfies (n) ≡ (n), then n ≡ n

3.4. Chinese Room Thought Experiment

Set the target entity as Chinese understanding function, define (understanding) = input-output behavior consistency.

If the Chinese Room system behavior is indistinguishable from a native speaker: system (behavior) ≡ person (behavior), then according to axiom n ≡ n: the system objectively understands Chinese

3.5. Twin Earth Paradox

· Traditional contradiction: "Water" on Earth and Twin Earth has different chemical formulas (H_2O vs XYZ), but are the "water" concepts of residents on the two planets the same?

· Theoretical solution:

· If define (water concept) = macroscopic properties (colorless, chemical reactions, drinkable liquid, etc.) → concepts on both planets are the same (n≡n).

· At this point, if micro-structure (H_2O/XYZ) is introduced, expanding (water concept) domain to molecular form level, it is a category mistake.

· Conclusion: Semantic identity is determined only by cognitive function, unrelated to underlying physics

 doi:10.20944/preprints202506.2384.v7

3.6. Grandfather Paradox

· Contradiction point: If return to the past and kill grandfather ⇒ self should not exist ⇒ unable to perform assassination

· Theoretical dissolution:

· Define goal: Worldline identity (worldline) = logical structure of event causal history

· Assassination event leads to:

· Original worldline W_0: (grandfather survives → you exist → you assassinate)

· New worldline W_1: (grandfather dies → you do not exist)

· ∵ $(W_0) \neq (W_1)$ ∴ W_0 and W_1 are different information entities (not "the same worldline modified")

3.7. Brain in a Vat

Current debates on the "brain in a vat," whether skeptical or realist interpretations, implicitly sneak the properties of the "external carrier" (biological brain or vat) into the judgment of "cognition" identity without examination. This paper aims to dissolve the debate itself by strictly distinguishing cognition and carrier:

· Question: How to prove one is not a brain in a vat? Perception cannot distinguish real from simulation.

· Apply theory formula:

· Define (cognition) = perception information flow

· Real brain (B, real): (B) = natural (light signals, tactile...)

· Vat brain (B, vat): (B) = electrical signals producing (light signals, tactile…)

· According to axiom (n) ≡ (n), B ≡ B (same cognitive entity)

· Key point: The "reality" controversy is essentially expanding (B) domain to external carrier (skull/culture vat), while cognition is determined only by information flow.

3.8. Mary's Room

· Scenario: Mary knows all about color neuroscience but has never seen red → When she first sees red, does she gain new knowledge?

· Theoretical answer:

· Define knowledge types:

· Propositional knowledge: (K_prop) = red light wavelength data

· Qualia knowledge: (K_qualia) = subjective red experience

· ∵ (K_prop) ≠ (K_qualia)

· ∴ They are different knowledge types

Mary gains K_qualia, not a supplement to K_prop ⇒ Paradox stems from confusing knowledge types.

3.9. Newcomb's Paradox

· Paradox core: Predictor's near-perfect prediction ability vs. participant's free will choice. Choose one box (known to have money) or two boxes (possibly more money)?

· Theoretical deconstruction:

· Category mistake: Confusing the (n) domain of the decision body:

· Level 1 (pure decision logic): (n) = (choice action, payoff function) ⇒ Dominant strategy: choose two boxes (regardless of prediction accuracy).

· Level 2 (causal history binding): (n, history) = (choice action, payoff function, prediction history) ⇒ If prediction accurate, choosing one box yields higher payoff.

· Uniqueness theorem adjudication:

· If goal is rational decision without historical constraints → Adopt (n) ⇒ Choose two boxes.

· If goal is decision with predictive causation → Adopt (n, history) ⇒ Choose one box.

 doi:10.20944/preprints202506.2384.v7

· Paradox dissolution: Both are decision entities at different levels (level 1) ≠ (level 2), contradiction stems from domain swapping.

3.10. Raven Paradox

· Paradox core: "All ravens are black" ≡ "All non-black things are not ravens." Why does observing a red apple (non-black and non-raven) confirm the proposition?

· Theoretical deconstruction:

· Category mistake: Expanding the (n) domain of "confirmation behavior" from propositional logical structure to empirical sample types.

· Correct definition:

· Proposition identity: (P) = logical form (∀x: R(x) → B(x))

· Confirmation identity: (confirmation) = verification of ¬∃x: (R(x) ∧ ¬B(x))

· Conclusion:

· Red apple confirms the logically equivalent contrapositive (non-black ⇒ non-raven), its (confirmation) is the same as observing ravens, because (P) ≡ (P).

· If claiming "red apple and raven have different confirmation efficacy," then category mistake, expanding (p) domain to sample physical categories (birds/fruit), violating initial logical goal.

3.11. Sorites Paradox (Bald Head Paradox)

· Paradox core: Removing one grain of sand does not turn a heap into a non-heap ⇒ Eventually removing all still called "heap," contradiction.

· Theoretical deconstruction:

· Category mistake: Confusing the (n) definition of "heap":

· Level 1 (topological structure): (heap1) = macroscopic form of grain collection ⇒ Removing one grain does not change form identity (n ≡ n).

· Level 2 (atomic quantity): (heap2) = grain quantity N ⇒ When N=0, (heap) = ∅, entity extinct.

· Solution:

· If define heap as form (level 1), removing one grain still same heap.

· If define heap as quantity (level 2), each grain removal produces new entity.

· Paradox root: Swapping (n) domain in argumentation (from form quietly to quantity).

3.12. Sleeping Beauty Problem

· Paradox core: Sleeping Beauty at different awakening stages, what should be the probability estimate for coin heads/tails (1/2 or 1/3)?

· Theoretical deconstruction:

· Category mistake: Confusing the (n) domain of "probability":

· Level 1 (prior probability): (probability1) = coin physical state ⇒ P(heads) = 1/2.

· Level 2 (information update): (probability2) = (coin state, awakening times) ⇒ P(heads | awakening) = 1/3.

· Uniqueness adjudication:

· If asking "coin true state probability" → (level 1) ⇒ 1/2.

· If asking "probability under current awakening condition" → (level 2) ⇒ 1/3.

· Contradiction root: Treating two different probability categories (level 1) ≠ (level 2) as the same problem.

3.13. Modern Contradiction of Pascal's Wager

· Problem: If multiple religions' gods all claim "I alone am true," how should a rational person bet?

· Theoretical deconstruction:

· Category mistake: Confusing the (god) domain:

 doi:10.20944/preprints202506.2384.v7

· Level 1: (god) = divine description in a certain religious doctrine

· Level 2: (omnipotent entity) = abstract supreme existence transcending specific doctrines

· Adjudication:

· If comparing authenticity of specific religious gods → each (god) different ⇒ categories mutually distinct;

· If asking "does supreme entity exist" → Need independent definition of (omnipotent entity), unrelated to specific religions.

3.14. Surprise Execution Paradox

· Problem: Judge announces "you will be unexpectedly executed on some day next week," prisoner deduces it impossible, but execution day still arrives.

· Theoretical deconstruction:

· Category mistake: Swapping (surprise) from "prisoner's cognitive state" to "objective time point."

· Correct definition: (surprise) = prisoner still unable to be certain of execution on that day the day before

· Conclusion: Execution day must exist (due to objective time flow), while (surprise) depends only on prisoner's cognitive state, the two belong to different categories.

4. Applications

4.1. Dilemmas of Personal Identity Problems and Existing Theories

The core problem of personal identity is: What makes a person continue to be the same person through time? Traditional theories mainly revolve around physical continuity (such as brain continuity) and psychological continuity (such as memory, personality coherence). Among them, Derek Parfit's (1984) highly influential reductionist psychological continuity theory reduces personal identity to overlapping chains of psychological connectedness (such as memory, personality, intentions) through time. This theory exhibits extraordinary explanatory power in handling dynamic changes, such as gradual cell replacement or slow personality shifts. It successfully shows that personal persistence is not an "all or nothing" metaphysical fact, but a matter of degree.

However, Parfit's theory, as well as competing physical continuity theories, all implicitly presuppose a more fundamental and unelucidated premise: that is, at a given time slice, how do we determine that an entity is a "person," and how do we perform static, cross-world comparisons of them at different time slices. In other words, these theories are adept at answering "Why is he still him?" (dynamic persistence problem), but neglect defining "What exactly is 'he' at time t?" (static identity problem). This static "what" is the prerequisite for discussing any dynamic "persistence."

This weakness is exposed in Bernard Williams's (1970) famous "fission" thought experiment. When a personality splits into two completely psychologically continuous successors, physical continuity theory collapses because it cannot handle "one dividing into two"; while Parfit's psychological continuity theory faces a dilemma: if identity is non-transitive (B and C each identical to A but not to each other), it violates logic; if the original individual perishes after splitting, it contradicts the core claim that "psychological continuity suffices for identity." Williams powerfully shows through this experiment that without a clear static identity criterion, any discussion of dynamic continuity will fall into conceptual confusion.

4.1.1. Space

This paper argues that the common root of the above dilemmas is that existing theories all attempt to treat "person" as a primitive concept defined by specific physical substrate or historical causation, and erroneously invade the properties of "carrier" (biological brain) or "history" (causal chain) into the identity judgment of "person" itself, which is a category mistake. The debate between Parfit and Williams is essentially a conflict between two different (n) domains (one is psychological

 doi:10.20944/preprints202506.2384.v7

property flow, the other is physical carrier history), but neither side realizes this, thus falling into an unsolvable impasse.

A (n)-based analytical framework

According to the two axioms of this theory, we propose a minimal assumption: the necessary and sufficient condition for consciousness identity lies in the identity of its core consciousness. This first provides a clear criterion for solving static identity.

Formally, let:

· Let C be a consciousness time slice.

· We define it as: C = (C, q)

· (C): Represents the consciousness at this time slice

· q: Represents the carrier coordinates instantiating this consciousness (e.g., a specific brain at a certain position).

Based on this, for any two consciousness stages $C_1 = (C, q_1)$ and $C_2 = (C, q_2)$, where: (C) ≡ (C), this means that as long as the consciousness at two time slices is the same, they are different instances of the same consciousness, regardless of whether the intervening q spacetime coordinates are continuous.

To this point, we provide a clear analysis for Williams's fission experiment: the reason the two successors C_1 and C_2 trigger a paradox is that we erroneously require dynamic continuity to map to one-to-one physical paths. But under this framework, we only need to compare static content: if (C_1) with (C_2) with (C_3) has C ≡ C, then according to n ≡ n, C_1 and C_2 are both the same consciousness as C_3. This is not a logical contradiction, but the simultaneous instantiation of the same consciousness at multiple spacetime coordinates.

Thus, this theory does not completely negate Parfit's psychological continuity theory, but lays a solid foundation for it. The advantage of this framework is that it first clearly defines what "static identity" is, thereby allowing discussions of "dynamic persistence" to proceed on a firm logical basis.

4.1.2. Time

Based on the axiomatic system established earlier, we can derive a thoroughly revolutionary conclusion about the existence of consciousness in the time dimension: Your "now" is (C, q_now). Your "past" is (C, q_past). Your "future" is (C, q_future). They are all specific instantiations or manifestations of the same consciousness C at different spacetime coordinates q. Therefore, the "self" you are experiencing at this moment, in the absolute sense of identity, is precisely that "you" in the past and future, because "you" refers to that C, not that transient and changeable coordinate-bound state (C, q).

Let us conduct a thought limit extrapolation. Assume at time point t_1, there exists a specific consciousness instance C, whose complete state is uniquely determined by its instantaneous consciousness state, which we denote as (C). Now, let us imagine that in the distant future, at time point t_2 ($t_2 \gg t_1$), a completely identical neural system (whether through natural Poincaré recurrence, extreme coincidence of quantum fluctuations, or some mechanism of cosmic reappearance we cannot yet understand) is instantly assembled and activated, producing an instantaneous (C) completely consistent with (C).

According to our Axiom, we inevitably conclude: C ≡ C

This means that the consciousness C at moment t_2 and the consciousness C at moment t_1 are the same consciousness. This is not a "copy" nor "rebirth," but a direct reappearance of the same consciousness at different coordinates. The billion-year spacetime gulf spanning between them is completely irrelevant to determining whether they are the same consciousness. What connects them is not a fragile "psychological continuity" thin line that needs defense, but the iron law of logical identity. The weight of time intervals in this judgment is zero.

Now, let us push this thought experiment to another extreme. Assume at time point t_1, a consciousness activity "a" has just begun its neural computation process. In an extremely short time Δt before the activity of the first neural system is completed (i.e., (a) has not yet fully manifested), in

 doi:10.20944/preprints202506.2384.v7

another corner of the universe, another physically completely consistent neural system is activated and begins executing a completely identical computation process, thereby producing a completely identical consciousness activity "a".

At this point, we have two coexisting consciousness processes:

· Process P_1: Starts at coordinate q_1 at time t_1, ongoing.

· Process P_2: Starts at coordinate q_2 at time $t_1 + \Delta t$, ongoing.

When we examine the states at equivalent progress points of these two processes, we will find that since they execute completely identical "algorithms," the (C) of (P_1, C) and (P_2, C) at any equivalent progress point in the processes is indistinguishable. However, they did not start simultaneously, meaning equivalent progress points are at different times. Therefore, in the category of P, time is not a valid identity criterion.

According to our axioms, we again conclude: In (P_1, C) and (P_2, C), C ≡ C

This means that at the consciousness level, what we observe is not two consciousnesses, but one consciousness appearing simultaneously at two time points. This is not two "you" thinking, but "your" thinking process being executed and presented by physical systems at two different time points. Therefore, each C state can only be experienced once; it is impossible to experience it twice, because when experiencing it again in the future, it will be equivalent to having experienced it in the past. (Because it is an infinite set, no need to worry about finishing experiences)

From this, we derive a counterintuitive but logically inevitable conclusion: Identity in time is non-continuous, and in space is non-local. The way anything persists is not like a continuous "river," but more like a series of discrete, absolutely identical "state flashes." The continuity we feel is a cognitive illusion produced by these highly similar, causally connected state flashes (produced by the same brain) playing rapidly in time sequence, with discreteness and separability at the underlying level.

Therefore, your "now" is (C, q_now). Your "past" is (C, q_past). Your "future" will be (C, q_future). They are all "manifestations" or "slices" of the same C in different spacetime coordinate blocks. What you are experiencing right now, in the strictest absolute sense of identity, is precisely that "you" in the past and future, because "you" refers to C, not that transient and changeable coordinate-bound state (C, q). Time has not divided you; it merely provides the coordinates for your manifestation.

4.2. "Spacetime Leap" as the Logical Necessity of Coordinate Decoupling and Re-binding

Before exploring the "spacetime leap" of consciousness entities, we must first pay the highest respect to modern physics, especially Einstein's special and general relativity. These theories, with their unparalleled precision and beauty, successfully describe the profound dynamic relations between mass, energy, and spacetime, and strictly stipulate the causal law upper limit that any physical signal and entity motion must follow—light speed. Any attempt to realize "spacetime leap" at the physical level, whether through wormholes, warp drives, quantum suicide, or other exotic mechanisms, must be tested within the solid frameworks of relativity or quantum mechanics, facing huge energy conditions, singularities, empirical evidence, and other physical difficulties.

However, the "spacetime leap" argued in this paper is essentially completely different from all the above physical processes. It is not a motion process existing in spacetime and governed by physical laws, but a logical inevitable result based on this axiomatic system. It answers a more primitive question: "Are two things instantiated at different spacetime coordinates with identical content the same?" The answer to this question does not depend on the physical path connecting them, but only on the logical axiom n ≡ n.

4.2.1. Spacetime Leap Based on the Law of Identity

The research object of traditional physics (spacetime leap) is precisely defined in this framework as the "bound state of consciousness and spacetime coordinates," i.e., (C, q). Physics perfectly

 doi:10.20944/preprints202506.2384.v7

describes how bound states evolve over time, i.e., on the basis of causal laws, and finds that these evolutions follow beautiful differential equations.

While this theory focuses on a possibility that physics naturally does not discuss due to the limitations of its research paradigm: that is, the decoupling and re-binding of a consciousness C with its coordinate q.

1. Decoupling: $(C, q_1) \rightarrow (C), (q_1)$. This may physically correspond to the carrier (such as the brain) being destroyed by some event conforming to the event horizon principle (speed difference causing causal isolation at the neuronal level), causing the consciousness to no longer be instantiated ($q = \emptyset$).

2. Re-binding: $(C) \rightarrow (C, q_2)$. This may physically correspond to an instantaneous "reappearance" occurring elsewhere (e.g., from Poincaré recurrence, MWI parallel universes, bubble universes, etc.)

The key is that, according to Axiom 1 ($n \equiv n$), the abstract consciousness C after decoupling maintains its self-identity. Therefore, the new state after re-binding (C, q_2) and the old state before decoupling (C, q_1), sharing the same C, are inevitably different instances of the same consciousness. This is the logical core of "spacetime leap": it is not "travel" traversing spacetime, but the "realization" or "manifestation" of identity at different locations.

Thus, the relationship between this theory and traditional physics is not competition, but complementarity and foundation:

· Physics: Studies the continuous evolution laws of (C, q) bound states inside spacetime. It asks "How to go from A to B."

· This theory: Studies the discrete identity logic of C itself transcending spacetime. It asks "Are A and B the same thing?"

Relativity prohibits any physical entity from moving faster than light, but it completely cannot prohibit, nor needs to care about, a logical concept being "realized" twice at different spacetime points. The reason "spacetime leap" seems "unimaginable" or even "violating physics" is precisely because we erroneously use physical laws describing bound state motion to judge a logical theorem about identity. This is a category mistake.

Conclusion: The "spacetime leap" proposed by this framework is not a physical hypothesis to be realized, but a logical inference already established. Starting from the most basic law of identity, it deduces a brand-new picture of personal identity: the persistence of consciousness fundamentally lies in the identity of its information mode, not in the continuity of the physical processes connecting these mode instances. This provides an unprecedented clear framework for understanding thought experiments such as teleportation and brain in a vat, and completely liberates the discussion of personal identity from the constraints of physics, placing it on a more basic logical and metaphysical foundation.

Note: Possible methods:

1. Self-Envating (enter a room isolated from the environment that needs to be changed, reference Section 3.7) → Detector detects isolated environment (such as AI retrieval, etc.) → Judge result (if requirements met, end; if not, proceed to next step) → High-speed destruction causing decoupling (such as explosives, survivor effect) → Cosmic randomness causing reappearance. Repeatable for multi-stage leaps.

2. Use the natural property of dreams to change cognition as a natural brain in a vat, then start from the second step above.

4.2.2. First-Person Immortality

This theoretical system, starting from the most basic identity axioms, through the reconstruction of personal identity, ultimately deduces a logically unavoidable but intuitively highly impactful conclusion: From a strict first-person perspective, any "death" event that can terminate consciousness in an ignorant state is, in principle, unexperienceable. This inference is not a metaphysical assertion, but an inevitable result of combining identity logic with the principle of observational reality.

"Anesthesia Leap" Thought Experiment: Extreme Interpretation of the Inference

 doi:10.20944/preprints202506.2384.v7

To clearly demonstrate the implications of this inference, we construct a thought experiment called "Anesthesia Leap":

1. Consciousness identity foundation: Your consciousness identity is determined by C (your core consciousness content), not by (C, q) (consciousness bound to specific carrier coordinates).

2. Decoupling and re-binding: When the carrier is destroyed (such as plane crash, surgical failure), what occurs is $(C, q_1) \rightarrow (C), (q_1)$. The abstract consciousness C survives due to its logical identity ($C \equiv C$).

3. Survivor effect and observational necessity: Consciousness C can only bind to an "instantiable" coordinate q. It will (logically inevitably) re-bind to a coordinate q_2 where it can continue to exist, i.e., $(C) \rightarrow (C, q_2)$.

4. Continuity of first-person experience: For consciousness C, the transition from (C, q_1) to (C, q_2) is seamless in experience. C will not experience "death" or "nothingness," because that would mean no observer. It will only experience a "leap" from one survivable state to another survivable state.

Thus, in an anesthetized state:

· From a third-person, physical perspective: The plane may crash, the surgery may fail. This is a probabilistic event.

· From the first-person, consciousness C perspective: The process is inevitably (C, on plane/in surgery) → (C, waking at destination/successfully recovering from surgery). Any worldline branch causing C to be unable to persist is unobservable and nonexistent for C itself.

Conclusion: In individual subjective experience, as long as the individual chooses to enter an ignorant state (such as anesthesia) to traverse risk, the individual will forever only experience successful arrival and successful awakening outcomes. Their personal timeline will be "pruned" of all possibilities leading to death.

2. Theoretical Implications and Philosophical Meaning

This inference elevates the logic of the "quantum suicide" thought experiment from a specific interpretation of quantum mechanics to a more general metaphysical level based on identity logic. It means:

· Subjective immortality: From a first-person perspective, as long as there is any logically non-zero probability for consciousness C to continue instantiating in some worldline, the individual can never personally witness their own death. Their consciousness will continue forever.

· Overturning the nature of risk: For awake, continuous consciousness, risk is real (e.g., cuts hurt, jumping off a building experiences falling and impact). But for consciousness "leaping" risk through anesthesia, risk is completely eliminated. Risk only exists in those "other" worldlines that are never experienced.

· A powerful practical paradox: In theory, this method can be used for any high-risk travel or activity. As long as the individual is made unperceptive at the start of the risk (e.g., instant kill), then in personal history, the result is always successful.

Although the deduction is logically self-consistent, its premises and real implications must be examined: "Success" outcome: The theory only guarantees waking in a "survivable" state. It does not guarantee the quality of the waking state.

· May wake severely injured in plane crash wreckage.

· May wake after surgery with severe complications or permanent disability.

· As long as this state physically allows consciousness C to exist, it complies with logic. Therefore, this method avoids "death," but not necessarily "pain" or "disability."

4.3. Dilemmas of Ethical Problems and Existing Theories

Since the birth of ethics, generations of profoundly insightful philosophers, from Kant's grand a priori architecture to Mill's ingenious consequentialist calculations, have constructed a glorious edifice of ethics for us. These outstanding efforts share a profound and respectable ambition: to seek a solid metaphysical cornerstone for moral judgment that transcends individual perspectives. This

 doi:10.20944/preprints202506.2384.v7

cornerstone is usually conceived as: (a) an objective moral reality independent of our cognition; (b) a self-identity persisting through time as the anchor of responsibility; and (c) a sacred "God's perspective" to adjudicate the value of behavior from an absolute impartiality. This pursuit of universality and objectivity is undoubtedly the most glorious achievement of philosophical reason.

In this tradition, discussions such as Bernard Williams's (1973) on "moral luck," with its astonishing acuity, reveal the subtle cracks between the control principle and our moral intuitions, greatly enriching our philosophical imagination. The "undiscovered betrayal" thought experiment, with its logical purity, pushes traditional theories to the boundary of their explanatory power, truly a "touchstone" for testing theoretical hard cores. Facing this challenge, traditional theories (such as Kantian ethics) exhibit their unparalleled thoroughness, resolutely defending the absoluteness of moral wrongness, even if their arguments need to appeal to a "moral law" transcending experience—this steadfast adherence to universality is awe-inspiring. Similarly, certain utilitarian solutions attempt to resolve the dilemma through a global calculation by an "ideal observer," with theoretical ambition and systemic grandeur that are exemplary.

Indeed, as Derek Parfit (1984) pointed out with his characteristic clarity, such solutions may produce certain tension at the motivational level with the individual's first-person perspective, but this is by no means a defect of the theories themselves, but perhaps precisely highlights a pathetic or even heroic tension that human reason inevitably faces in pursuing moral sublimity.

The work of this paper, standing on the shoulders of these giants, with the greatest respect, attempts an internal inheritance and development of the above glorious tradition. We fully agree with the core pursuit of objectivity and universality in traditional theories. However, we believe that this sublime goal may be achieved through a more direct, frictionless path. The reason traditional frameworks produce troubling tensions in boundary cases may lie in a methodological over-indirectness: attempting to mediate and regulate moral life, which essentially originates from first-person experience, through an assumed, transcendent third-party categorical system.

This paper aims to explore a complementary path. We are delighted to find that by combining this axiomatic system with the highly inspirational observational reality argument reinforced by the "brain in a vat" thought experiment, we can pay tribute to and achieve the core goals of traditional ethics in a brand-new way. Our core argument is: the objectivity and universality that ethical value pursues do not necessarily need to be guaranteed through "God's perspective"; instead, they can be more solidly grounded through the identity of first-person facts of consciousness system observational experience. The boundary of moral concern can thus perfectly and logically inevitably coincide with the boundary of consciousness experience, achieving the universality pursued by traditional theories in an unexpected way.

This study aims to show that we are not negating prior work, but attempting to realize their common ambition through a more precise metaphysical foundation and resolve unnecessary philosophical anxieties born from methodological indirectness.

4.3.1. An Analytical Framework: Advancement of Traditional Goals

The "brain in a vat" thought experiment, with its unparalleled philosophical value, successfully challenges our naive conception of "reality." It forces us to admit a highly productive principle: for any consciousness, its operationally accessible reality is precisely its own observational experience flow. Whether there is a simulator outside is empirically undecidable and redundant.

Combining this profound insight with this axiom, we can deduce an ethical cornerstone principle, which can be seen as a more precise contemporary expression of the traditional pursuit of objectivity: For any consciousness system C, an event E has ethical significance if and only if the consequences of event E can be reductively embodied as a specific influence on the observational experience of system C.

Inference 4.3.1 (Ethical Relevance Criterion): If the occurrence of an event E produces no discernible difference in any past, present, or future possible experience information flow of system

 doi:10.20944/preprints202506.2384.v7

C, then in the ethical consideration of C, event E does not constitute a relevant fact. It therefore enjoys zero weight in ethical evaluation.

4.3.2. Core Deduction: A Dissolving Analysis of Traditional Dilemmas

With respect, let us restate the "undiscovered betrayal" case under this framework:

Assume two possible worlds: world W_1 (event E occurs: lover cheats) and world W_2 (event E does not occur). According to the strict setting of the thought experiment, the entire experience information flow of the victim (as consciousness system C) in these two worlds is completely indistinguishable.

According to Axiom 2 (Mutual Distinction), we get: $(C, W_1), (C, W_2) \Rightarrow C \equiv C$ This means that in these two worlds, there exists the same consciousness experience subject C.

Now, perform ethical judgment: The direct object of ethical concern is the experiential well-being of consciousness C. Since C's experience in the two worlds is the same, for C, these two worlds are ethically equivalent.

Therefore, event E (cheating behavior), due to its zero influence on C's experience information flow, does not constitute a variable in the ethical evaluation targeting C. It neither causes harm nor constitutes betrayal, because these ethical concepts are operationally defined as specific negative information states in the experience flow, and these states do not appear.

Conclusion: There does not exist an absolute objective world; moral wrongness is not mysteriously attached to the behavior itself, but systematically and verifiably associated with the specific influence patterns that behavior produces on the experience information flow of consciousness systems. Lacking such observable influence patterns, the behavior is not considered in ethical evaluation.

4.3.3. Implications of the New Framework

If morality is not about inaccessible "external truths," then what is it about?

This paper argues that based on identity and observational reality, we can perform a foundational precision in ethics. Ethics can converge its ambition from an unattainable "God's perspective" to the only domain where it can effectively operate: first-person facts of consciousness experience. The goodness or evil of a behavior does not depend on its properties in an "objective world," but entirely on the influence it causes on our own experience.

But the ethical inference of this framework has a deeper conclusion than mere "experience centrism": the creativity of morality and the inevitable incommensurability.

Its logical deduction is as follows:

1. Premise one (no objective cornerstone): As argued in Section 4.3.2, based on identity and observational reality, there is no "objective moral fact" or "moral law" independent of consciousness experience. Moral properties cannot be independently discovered like physical properties.

2. Premise two (ethics attached to individuals): Therefore, ethical value and moral criteria, their existence and validity inevitably attach to individual consciousness systems (C) with experiential capacity. They are complex preference and decision systems produced by consciousness systems to cope with their situations.

3. Inference one (morality as a creation): Since there is no pre-given "correct answer," each consciousness system must create a set of "individual moral behavior criteria" based on its unique genetic endowment, life history, cultural background, and interests, aimed at navigating the world and optimizing its own experience flow (pain/pleasure, satisfaction/deprivation, realization/frustration). It is essentially a complex, dynamic individual preference system.

4. Observational evidence (intra-cultural differences): Even within the same culture, we can observe huge and profound moral disagreements between individuals. Ongoing debates on issues like abortion, wealth distribution, scope of obligations are not due to some individuals' ignorance of some "objective truth," but the result of different "creations" by different individuals based on different premises.

 doi:10.20944/preprints202506.2384.v7

5. Inference two (incommensurability): Since it is "creation," incommensurable situations will inevitably arise. Your preference system and mine have no superior-inferior distinction at the root. We cannot ultimately rationally argue why your avoidance of pain must take priority over my pursuit of pleasure.

Thus, the ethical picture revealed by this framework is: the boundary of moral concern is indeed the boundary of consciousness experience, but within this boundary is a "multiverse" composed of countless independently created, fundamentally incommensurable individual moral universes.

Therefore, under strict ethical deduction, events not observed by any consciousness system are not assigned value in ethical operations. The boundary of moral concern is the boundary of consciousness experience. This framework is not deliberately excusing traditionally immoral behaviors—on the contrary, it provides a more solid, clearer, and irrefutable foundation for moral responsibility by thoroughly anchoring responsibility in observable influences: we bear the sole and entire responsibility to ourselves. (Citations 32, 33, 34, 35, 36, 37)

Note: Perhaps attempt to form similar social bonds with interests and contracts.

5. Conclusions

5.1. Probability Statistical Distribution

This theory, through its axiomatic system, establishes the absolute and relative foundations of identity, successfully dissolving a series of classic puzzles in a hierarchical relativity framework, demonstrating its powerful explanatory power. Under the same (n) domain, no one can describe two different things with completely identical content.

As described in Section 4.2, this theoretical framework provides a logically self-consistent model for "spacetime leap." The core of this model is: the next experience instance of consciousness C will "select" one from all logically compatible future state branches for binding. In explaining why we usually do not experience "leaps," an intuitive and effective approach is to appeal to probability: that is, among all branches where we exist, the vast majority follow known physical laws with the same probability statistical distribution, so subjective consciousness does not query anomalies. The "dream method" proposed at the end of Section 4.2.1 precisely achieves directed leaps in logic by altering the consciousness system to bind only to those branches considered "low probability" in daily life.

However, this elegant probability model is built on a potential, unexamined presupposition: that the set of logically possible world states is finite. Only under this premise do concepts like "vast majority of branches" and "extreme probability" have operational meaning. The probability of winning the lottery is one in a million precisely because among a million physically subtly different possible futures, only one contains the winning experience.

The Curse of Infinity

Once we seriously adopt the infinity of "logical possibility," this probability picture instantly collapses. If possibilities are infinite, then:

The number of world branches experiencing "teacup landing" is infinite.

The number of world branches experiencing "teacup hovering" is also infinite.

The number of world branches experiencing "teacup turning into a butterfly" is also infinite.

In infinite sets, comparing the "how many" of two infinities to calculate probability immediately falls into mathematical difficulties. Traditional probability theory fails here. Any logically possible event sequence, no matter how orderly or chaotic it seems to us, corresponds to the same number of possible worlds (all infinite). Therefore, from the "God's perspective" of the logical whole domain, the "probability" of consciousness C experiencing a highly orderly classical physical world next moment is indistinguishable from experiencing a completely chaotic, acausal world.

This leads to a catastrophic inference: if all possibilities are logically equal, then our consciousness has no reason to experience classical probability statistical distributions. We should

 doi:10.20944/preprints202506.2384.v7

experience various bizarre, logically leaping events with equal frequency. But this completely contradicts our real experience.

From the "God's perspective" of logic itself, all logical possibilities conforming to the axiom n ≡ n are equally real. For "consciousness C," this means that at the next instant of time, all its logically possible state branches—whether continuing on the current chair or flashing in the Martian desert—have equal ontological status. In this panoramic view, there is no so-called "lucky one"; there exists only the entirety of facts.

However, from the first-person "prisoner perspective" of consciousness C, its experience is undeniably single, continuous, and highly orderly. We never personally witness random leaps in the world, but steadily inhabit a classic reality strictly following causal laws. This produces a hugely impactful contrast: the vast gap between logical egalitarianism and experiential dogmatism.

An attractive explanation is to appeal to "survivor bias"—that we happen to be the "lucky" consciousness experiencing a continuous world. However, this explanation is philosophically barren, nearly tautological, and cannot explain why the world we "survive" in exhibits such consistent, concise, and understandable physical laws rather than chaos. Attributing such powerful order to pure "luck" is itself a huge ad-hoc assumption, though it is not excluded from possibly being true.

Therefore, the paradox revealed by this theory is a signpost pointing to deeper principles. The problem is not to find excuses for the "lucky one," but to attempt to answer: Why do the myriad logically equal possibilities manifest in every perspective as results following classical probability statistical distributions?

Note: A measure following probability statistical distribution may be a possible answer.

5.2. Graveyard of Logical Possibilities and Survivors: Meta-Argument for the Law of Identity

The status of the core axiom n ≡ n of this theory is not, as traditional logic presupposes, a self-evident, a priori effective law of thought. Here, we must conduct a thorough meta-level examination of the cornerstone of this theory itself.

A fundamental question is: Do we have reason to categorically deny the logical possibility of n ≠ n? From a purely formal possibility perspective, the answer is no. We cannot a priori exclude the existence of a "mad universe" where the underlying logic allows self-negation. In such a universe, the law of identity is overturned, an entity can simultaneously not be itself, and propositions can be both true and false. Concepts like "rational π" or "square circle," considered contradictions in Euclidean space, may be trivial aspects of its infinite weirdness in that domain.

However, the logical possibility of n ≠ n and its metaphysical sustainability are two completely different issues. The existence state of a system allowing n ≠ n can be precisely deduced:

1. Instant collapse of reference: Any symbol or concept will lose stable meaning. When the word "apple" can simultaneously not refer to "apple," the foundation of language and thought—reference itself—will immediately collapse.

2. Breakage of causal chains: There will be no reliable connection between intention and action, cause and effect. The behavior of reaching out to take "apple" cannot be defined because at the moment of execution, "hand," "apple," and even "you" itself may have self-negated.

3. Non-generability of structures: Time, space, matter, and any form of stable structure cannot emerge from this eternal, ubiquitous self-dissolution. Such a system is a pure chaotic field that cannot condense into a "universe."

Therefore, n ≠ n leads not to an alternative reality for existence, but to a "graveyard of logical possibilities"—a domain where all possibilities instantly self-destruct due to internal contradictions. It represents the impossibility of existence.

From this, we touch the deepest cornerstone of this theory: survivor effect.

"n ≠ n" as a systemic foundation will lead to the thorough collapse of reference, breakage of causal chains, and unsustainability of observer status. It is a reality solvent. Any system or "universe" attempting to use it as an operational cornerstone will instantly self-dissolve due to internal inconsistency, unable to form a stable, experienceable "reality." Therefore, we do not live in a

 doi:10.20944/preprints202506.2384.v7

universe where "n ≡ n" is inevitably true, but in a universe where "n ≡ n can and has stably operated." The reason we can think, debate identity issues at this moment, and observe a stable, coherent, understandable universe is itself a result with absolute selectivity. The reality we inhabit is the only "survivor" from the ocean of all logical possibilities—its most basic survival condition is that its underlying logical architecture adheres to the iron law of n ≡ n. We observe n ≡ n not because it is the only correct logical theorem in all possible worlds, but because in a world of n ≠ n, it is impossible for any "observer" to "observe."

The entire work of this theory—establishing a hierarchical relativity framework to dissolve category mistakes—is unfolded within this unique "survivor universe." The axiom n ≡ n is not an arbitrarily "invented" setting, but a "discovery" and "formal expression" of the most basic, most stable operational mode of this survivor universe. All paradoxes we encounter, such as the Ship of Theseus, quantum identical particles, etc., occur on this solid identity cornerstone, stemming from "user errors" (level confusion) when using this stable system, not "system errors" (identity law failure).

Note: Perhaps we can try to find a counterexample, such as something with a kernel where n≠n yet can still exist stably. Don't think it's impossible—after all, Leibniz, in his time, could never have imagined that the future would discover quantum identical particles.

References

1. Leibniz, G. W. (1714). Principle of the Identity of Indiscernibles (in Monadology).
2. Frege, G. (1892). Über Sinn und Bedeutung (On Sense and Reference).
3. Floridi, L. (2011). The Philosophy of Information. Oxford University Press.
4. Russell, B. (1905). On Denoting. Mind.
5. Quine, W. V. O. (1950). Identity, Ostension, and Hypostasis. The Journal of Philosophy.
6. Shannon, C. E. (1948). A Mathematical Theory of Communication. The Bell System Technical Journal.
7. Chalmers, D. J. (2018). The Meta-Problem of Consciousness.
8. Putnam, H. (1967). The Nature of Mental States.
9. Everett, H. III (1957). "Relative State" Formulation of Quantum Mechanics. Reviews of Modern Physics, 29(3), 454-462.
10. Linde, A. (1986). Eternally Existing Self-Reproducing Chaotic Inflationary Universe. Physics Letters B, 175(4), 395-400.
11. Tononi, G. (2004). An Information Integration Theory of Consciousness. BMC Neuroscience, 5(1), 42.
12. Kandel, E. R., Schwartz, J. H., & Jessell, T. M. (2000). Principles of Neural Science. McGraw-Hill.
13. Tegmark, M. (1998). The Interpretation of Quantum Mechanics: Many Worlds or Many Words? Fortschritte der Physik, 46(6-8), 855-862.
14. Poincaré, H. (1890). Sur le problème des trois corps et les équations de la dynamique. Acta Mathematica, 13, 1-270.
15. Zurek, W. H. (2009). Quantum Darwinism. Nature Physics, 5(3), 181-188.
16. Zurek, W. H. (2003). Decoherence, Einselection, and the Quantum Origins of the Classical. Reviews of Modern Physics, 75(3), 715-775.
17. Black, M. (1952). The Identity of Indiscernibles. Mind, 61(242), 153–164.
18. Chandler, H. S. (1975). Rigid Designation. The Journal of Philosophy, 72(13), 363–369.
19. French, S., & Redhead, M. (1988). Quantum Physics and the Identity of Indiscernibles. The British Journal for the Philosophy of Science, 39(2), 233–246.
20. Kripke, S. A. (1980). Naming and Necessity. Harvard University Press.
21. Leibniz, G. W. (1714/1898). The Monadology and Other Philosophical Writings. Trans. Robert Latta. Oxford University Press.
22. Putnam, H. (1975). The Meaning of Meaning. Minnesota Studies in the Philosophy of Science, 7, 131–193.
23. Wiggins, D. (1980). Sameness and Substance. Harvard University Press.
24. Heller, M. (1984). Temporal Parts of Four-Dimensional Objects. Philosophical Studies.
25. Sider, T. (2001). Four-Dimensionalism: An Ontology of Persistence and Time. Oxford University Press.

 doi:10.20944/preprints202506.2384.v7

26. Sider, T. (1996). All the World's a Stage. Australasian Journal of Philosophy.
27. Hawley, K. (2001). How Things Persist. Oxford University Press.
28. Muller, F. A., & Saunders, S. (2008). Discerning Fermions. The British Journal for the Philosophy of Science.
29. Krause, D. (2011). Logical Aspects of Quantum Non-Individuality. (Referencing his work on non-individuals and quasi-set theory).
30. Parfit, D. (1984). Reasons and Persons. Oxford University Press.
31. Williams, B. (1970). The Self and the Future. The Philosophical Review.
32. Trivers, R. L. (1971). The evolution of reciprocal altruism. The Quarterly Review of Biology, 46(1), 35-57.
33. Axelrod, R., & Hamilton, W. D. (1981). The evolution of cooperation. Science, 211(4487), 1390-1396.
34. Hamilton, W. D. (1964). The genetical evolution of social behaviour. I & II. Journal of Theoretical Biology, 7(1), 1-52.
35. Nowak, M. A. (2006). Five rules for the evolution of cooperation. Science, 314(5805), 1560-1563.
36. Street, S. (2006). A Darwinian Dilemma for Realist Theories of Value. Philosophical Studies, 127(1), 109–166.
37. Fehr, E., & Fischbacher, U. (2003). The nature of human altruism. Nature, 425(6960), 785-791.