

Article

Not peer-reviewed version

Steins Theory: A New Axiomatic System for Identity

[Jiaqi Guo](#) *

Posted Date: 9 January 2026

doi: 10.20944/preprints202506.2384.v8

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



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a [Creative Commons CC BY 4.0 license](#), which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

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 the externalist position to its extreme, successfully challenging the internalist model of meaning and setting the basic agenda for debates on referential determinacy for the subsequent decades. However, despite the highly inspirational nature of these pioneering works, an intriguing phenomenon is that the debates they sparked—such as discussions around core cases like Theseus’s ship and identical particles—seem to have fallen into a kind of impasse. This article attempts to argue that this impasse may not stem from the depth of the problem itself, but precisely from an unexamined deep presupposition shared by these otherwise highly convincing theories: namely, the belief that there exists some single, decisive level (whether microscopic physical structure or historical causation) that can once and for all answer the identity question. This article proposes that, rather than continuing to seek a superior single answer under this presupposition, a more productive approach may be to reflect on this presupposition itself. To this end, we develop an analytical framework of hierarchical relativity. Interestingly, this framework shows that those seemingly opposing excellent 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. Therefore, this framework is not intended to negate the work of predecessors, but aims to provide a new path for resolving a series of philosophical puzzles arising from category mistakes by clarifying the valid scope of application of those works.

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

1. Introduction

The problem of identity, namely “what is what it is,” is at the core of metaphysics and logic. From Leibniz’s (Gottfried Wilhelm Leibniz, 1714) highly insightful “Principle of the Identity of Indiscernibles” (PII)—which set a lofty ideal for the individuation of entities with its logical simplicity and power—to Kripke’s (Saul Kripke, 1980) pioneering theory of necessary identity based on rigid designators and origins, which provided a seemingly solid foundation for the stability of reference with its profound grasp of modal intuitions. The excellent efforts of generations of philosophers have collectively built a grand intellectual tradition for our understanding of individual persistence and identification.

However, an intriguing phenomenon is that these theories, which are highly convincing within their respective domains, often exhibit a regrettable, systematic limitation when dealing with complex boundary cases posed by reality and thought experiments. Leibniz’s strong PII principle encounters fundamental difficulties in the face of quantum identical particles that perhaps were not foreseen in its original conception (French & Redhead, 1988). Its inherent rigor paradoxically leads 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 ingenious theory, aimed at anchoring reference rigidly, becomes quite entangled in its explanatory power when handling diachronic changes in intrinsic properties (such as Theseus’s ship) (Chandler,

1975), let alone its difficulty in accommodating the identical particle phenomena presented by quantum mechanics that challenge the classical view of individuals.

Thus, we face a peculiar intellectual impasse: whether Leibniz’s strong program pursuing property identity or Kripke’s historical path focusing on the necessity of origins, they seem to successfully illuminate one wing of the identity edifice but unfortunately plunge the other side into deeper shadows. Their shared, perhaps unspoken ambition—to find a single, absolute criterion for identity—itself, does it constitute an a priori obstacle that prevents us from truly understanding the multidimensional essence of the problem?

This article believes 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 article 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: namely, mistakenly attempting to use an answer from one level to respond to a question from another level.

This article aims to transform the classical controversies in ontology and semantics into a clear, operable conceptual choice problem. Thereby, not to solve these puzzles, but to dissolve them, providing a new way out for a series of philosophical anxieties arising from category mistakes.

Note: In this article, unless propositions have explicit premises (such as specifying consideration from a certain perspective), the default premise for propositions is that they are effective in reality (physics). This default is because the article considers that if a proposition does not pursue real-world effectiveness, then the default shift in perspective naturally makes the truth or falsity of the proposition’s conclusion unimportant.

2. Analysis

2.1.1. Axiom 1 (Self-Identity)

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

2.1.2. Axiom 2 (Mutual Distinctness)

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

Core Corollary:

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

Note: See example in Section 3.1

2.2. Category Mistakes and Hierarchical Confusion: Taking Theseus’s Ship as an Example

In dealing with the enduring puzzle of Theseus’s ship, a highly influential and intuitively appealing scheme has been proposed, with representatives including David Wiggins (1980). This scheme argues that the identity of an object is not guaranteed by its instantaneous properties at any moment, but must be carried by its spatiotemporal continuity and an uninterrupted historical causal path. The great advantage of this approach lies in its successful capture of our deep intuition about the concept of “objects persisting in time”—namely, that things are not instantaneous existences, but have their life history (career) or biography.

Another scheme, namely four-dimensionalism (Perdurantism), provides a completely different and highly metaphysically concise picture (see Heller, 1984; Sider, 2001). This theory, with its thorough clarity, extremely cleverly avoids many traps of diachronic identity. Four-dimensionalism argues that Theseus’s ship is not a three-dimensional entity that “fully exists” in time, but a “spacetime worm” extending in four-dimensional spacetime. Each time-slice of the ship is regarded as a temporal part of this four-dimensional object. Thus, the so-called “change” merely means that this four-dimensional whole possesses different properties (such as different planks) on different

temporal parts. Under this framework, the initial Theseus's ship (as a four-dimensional entity), the replaced ship, and the ship reassembled from old planks are three different four-dimensional objects. They may have completely identical three-dimensional cross-sections at some time slice (thus indistinguishable at that moment), but as wholes, they are naturally different. The excellence of this scheme lies in converting the problem of persistence from the troubling "identity" to the relatively clear "part-whole" relation.

Closely related to four-dimensionalism is stage theory (see Sider, 1996; Hawley, 2001), which, while retaining many advantages of four-dimensionalism, attempts to better accommodate our everyday linguistic intuition that "objects are three-dimensional." Stage theorists believe that what we usually call "Theseus's ship" 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 "the ship is the same" at time t , we are actually saying that there exists a (primitive) counterpart relation between the stage at t and the stage at the previous t_1 , maintained by some similarity and causal continuity. Stage theory, with its conceptual economy, avoids presupposing identity relations across time, thereby exhibiting strong theoretical appeal.

However, despite the ingenious internal logic of the above theories and their ability to be self-consistent, this article 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 attempt to provide us with a single, absolute criterion for identity judgment. To achieve this goal, they must construct "time" or "spatiotemporal 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 and t_2 the same?" logically depends on verifying spatiotemporal coordinates or cross-temporal associations.

This article believes that it is precisely this key theoretical move that inadvertently leads to a "category mistake." Let us formally reconstruct the schemes of each theory: it actually adopts a domain of definition as: (physical properties of the ship, historical causal path/four-dimensional whole/counterpart relation).

Let us apply $n \equiv n$ to Theseus's ship. 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 are concerned with its identity as an "objectively verifiable falsifiable ship," namely the category (physical ship), with relevant properties {for example, all particles and arrangement shapes of the ship}.

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

This article believes that the controversy here stems from the 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" {under (physical ship)} and "no" {under (physical ship, history)}—are not contradictory because they answer two different questions. Imposing the answer from (physical ship, history) on the question from (physical ship) constitutes a category mistake. The true value of the historical path theory lies in revealing the existence of "history" as an important category, but it mistakenly regards it as the only decisive category.

We can formally reconstruct it as follows:

1. Initial question: Judge whether the entity "ship" at time points t_1 and t_2 is the same.
2. Correct (n) domain of definition (according to the initial question): Should only include properties related to the material of the "ship," namely physical ship: (all particles and arrangement shapes of the ship)
3. Category mistakes in historical/four-dimensional/stage theories:

- Wiggins actually adopts history + physical ship (all particles and arrangement shapes of the ship, historical causal path).
- Four-dimensionalism actually adopts 4D + physical ship (all particles and arrangement shapes of the ship, spatiotemporal coordinates).
- Stage theory actually adopts stage + physical ship (all particles and arrangement shapes of the ship, counterpart relation).

The process leading to this can be formally expressed as:

- They attempt to answer the identity question based on (physical ship): “Is Ship at t_1 $\equiv$ Ship at t_2 ?”
- 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_ t_1) = (physical ship_ t_2).
- But since the historical path, spatiotemporal coordinates, or counterpart relation have changed, (physical ship_ t_1) $\neq$ (physical ship + history/spacetime/stage_ t_2).
- Thus, they conclude Ship_ $t_1 \neq$ Ship_ t_2 to answer the physical ship question.

Thus, we see an interesting situation: each theory effectively answers a question, but possibly not the one originally 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 for “Is the ship structurally the same?” This is like a judge, asked “Does the defendant’s behavior comply with Article X of the criminal law?”, giving a verdict after consulting the civil code. The conclusion may be coherent within its own system, but it has quietly switched the debate’s venue, which is a category mistake.

Therefore, the true contribution of the historical path theory perhaps lies in excellently revealing how various properties such as “history” and “spatiotemporal whole” influence our identity judgments as powerful explanatory properties. But its limitation lies in attempting to elevate such explanatory properties to a metaphysical necessary condition, thereby having to expand the evaluation criteria for the identity problem to maintain the integrity of its theory. The value of this theory lies in not needing to make such a difficult expansion, but by clarifying the levels of the problem, allowing different domains/properties to give valid, non-conflicting answers to different questions. Not to solve these puzzles, but to dissolve them.

2.3. Conservation

2.3.1. The Indistinguishability of Identical Particles in Quantum Mechanics Poses the Most Severe Challenge to Leibniz’s PII (Citations 1,2), Yet Provides a Natural Model with Physical Empirical Basis for This Theory.

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

Saunders’s scheme 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 a creative perspective for seeking the basis of individuation in relationality. The complexity of this approach and the extensive discussions it has provoked prove its profound philosophical value. However, it is precisely this technical complexity that exposes a potential cost underlying the scheme: its definition of “purely extensional relational properties,” though striving for precision, inevitably introduces considerable terminological ambiguity, such that its defenders must also carefully handle accusations of circular argumentation (Muller & Saunders, 2008). More centrally, the entire theoretical edifice of the scheme is built on an unsettling presupposition: namely, 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 become an ad-hoc patching to save a premise. When applied to states with indefinite particle numbers in quantum field theory, this strategy of continuously introducing new relational properties to save individuality becomes increasingly ad-hoc: it no longer resembles an elegant inference of the theory itself, but more like an increasingly costly price paid to maintain the theoretical premise (that individuality must exist).

Faced with the dilemmas of revisionism, another revolutionary scheme 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 on individual identity are using the wrong category from the beginning. They should be understood as “non-individuals” and need to be described using highly specialized mathematical tools such as quasi-set theory.

Krause’s scheme is notable for its conceptual thoroughness and consistency; it unreservedly embraces the most counterintuitive features of quantum mechanics, decisively breaking with our entire classical framework about objects and spatiotemporal localization. This resolute stance is undoubtedly clean and efficient in theory. However, the corresponding cost brought by this efficiency is that the concept of “non-individual” itself imposes a considerable explanatory burden metaphysically, requiring us to abandon a whole set of deeply ingrained intuitive understandings about what “a thing” means.

2.3.2. This Article’s Solution: A Hierarchical Relative Framework

The above two schemes 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 level are we asking the identity question?”

This article 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 spatiotemporal coordinates.

- When we inquire at the (particle) = o level, i.e., only comparing 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 (particle state) = (o, q) level, since q (such as position) must differ, hence $(P_1) \neq (P_2)$, so they are different particle states. This explains why we observe multiple scattering events in experiments.

Therefore, the confusion brought by quantum identical particles stems from mistakenly intruding the distinctness at the q (spatiotemporal 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 logical concepts) 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 scheme lies in: it absorbs the advantages of Krause’s scheme in acknowledging the peculiarities of quantum (by interpreting “non-individuality” as identity at the o level), while avoiding its radical metaphysical cost (we are still talking about “quanta,” just different categories); at the same time, it explains why Saunders’s strategy of introducing relational properties seems feasible in some cases (because he mistakenly takes properties at the q level as the individuating basis at the o level), yet fundamentally goes astray.

2.3.3. 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 spatiotemporal coordinates (such as position x , time t).

Formalization:

1. When the domain of p is $P = (o, q_1)$, an electron at a certain coordinate
2. Coordinate decoupling (destruction): $(o, q_1) \rightarrow (o), (q_1) \Rightarrow$ The particle degenerates into a pure eigenstate (o) , and since it lacks observable basis ($q = \emptyset$), it is unmeasurable. $\forall$ particle states (o, q_1) and (o, q_2) , where it can be found: $(o) \equiv (o)$ indicates:

- When two particles' eigen properties are indistinguishable ($o \equiv o$), regardless of how their spatiotemporal coordinates $q_1 \neq q_2$ differ, their particles are the same electron $e = (o)$ projected in different spacetimes

Physical Interpretation:

- Particle annihilation $\Rightarrow$ Set decoupling rather than annihilation $\Rightarrow e = (o)$ enters a free state
- Particle creation $\Rightarrow$ The same e binds to a new coordinate $q_2 \Rightarrow$ Observed as reappearance, example: Electron e disappears at position q_1 and appears at q_2 , actually the coordinate migration of electron $e = (q=-1e, m_e, s=1/2...): (e, q_1) \rightarrow (e) \rightarrow (e, q_2)$, its electron identity guaranteed by $n \equiv n$.

Direct Corollary: Conservation Theorem: Logic allows existence, neither annihilates nor updates

2.4. Symmetry

Max Black's (1952) symmetric universe thought experiment poses the most extreme challenge to Leibniz's strong PII principle. He envisions a universe containing 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 argues that this is a real scenario of "two" things, thereby refuting PII—namely, there is no property that can be used to distinguish them, but they are still numerically different entities.

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

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

- If we define (sphere) as the set of all traditional properties (intrinsic + relational), i.e., $(n) = \{\text{mass } M, \text{ shape spherical, ..., distance } X \text{ from a sphere}\}$, then according to the axiom, since $(n) \equiv (n)$, we inevitably conclude sphere $\equiv$ sphere. This seems to directly reach the PII conclusion that Black tries to refute.
- However, Black's intuition—"there are obviously two spheres"—is not without basis. This theory interprets it as a mental fixation. The observer reports "seeing two" because their perspective itself is embedded in this symmetric spatiotemporal coordinate system. This article believes that a fundamental misconception shared by Black and his commentators lies in assuming that the references of "sphere" and "sphere" necessarily correspond to two entities with independent spatiotemporal coordinates. This presupposition leads them to only choose between "abandoning PII" or "inventing new metaphysical concepts." The concept of "Coordinate Self-Reference" provides a third path for this dilemma. Formalization: For the entire symmetric system S , define: $(S) = \{ \text{there exists a sphere, its property set is } P, \text{ and the sphere is opposite to itself} \}$ This description seems complex, but simply put, (S) describes a single coordinate framework that allows "self-facing." Within this framework, the sphere being opposite to itself is not a grammatical error, but an accurate description of a singular coordinate topology. The visually presented "two" spheres are the projection 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, but the topological property of space itself)

- System S: (S) describes the state after a single sphere binds to a special self-facing coordinate topology structure: (sphere, R_self-facing).
- Paradox Dissolution: Black's error lies in mistakenly inferring the existence of two spheres (sphere_1, sphere_2) from the system state (sphere, R_self-facing). He confuses categories, using the descriptive result of (S) to answer the question about (single sphere). In fact, there never existed a second sphere; what exists is always one sphere, placed in a special coordinate topology that can produce "double image projection."

Therefore, this framework does not deny our intuition of "seeing two spheres," but provides a brand-new, more precise ontological explanation for this intuition: that is the perception of a single sphere in a self-referential coordinate topology. This successfully resolves the surface contradiction between PII and counting intuition, while avoiding introducing any ad-hoc individuating factors. Black's challenge not only fails to refute the law of identity, but through the hierarchical analysis of this framework, more profoundly reveals the dependence of "identity" judgments on the background framework.

3. Examples

3.1. Duplicate 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 $\rightarrow (n) = \text{text semantics}$, then $n \equiv n$;
- If the goal is document positional entity identity $\rightarrow (n) = (\text{text semantics}, \text{position})$, then $(\text{content}, \text{Loc}_A) \neq (\text{content}, \text{Loc}_B)$.
- Conclusion: Duplicates are the same information formed by different spatiotemporal coordinates, leading to observability.

3.2. Gibbs Paradox

Category Mistake:

- The goal should be particle type identity $\rightarrow (n) = (\text{mass}, \text{spin}, \dots)$
- Classical statistics privately expands to $(n) = (\text{intrinsic properties}, \text{fictional labels})$
Correction: (n) and $(n, \text{labels}) \Rightarrow (n) \equiv (n)$ Entropy increase error stems from wrongly choosing (n) domain (introducing labels)

3.3. Black Hole Information Paradox

Category Mistake: Binding the domain for internal properties (n) to spatiotemporal coordinates $(n) = (\text{information structure}, \text{black hole coordinates})$

Correct Solution:

- Define goal: Internal property identity $\rightarrow (n) = \text{quantum properties}$
- Black hole disassembles the set (quantum properties, coordinates), unpaired coordinates lead to unobservability, but (quantum properties/coordinates) as logical concepts do not disappear
- If new spacetime satisfies $(n) \equiv (n)$, then $n \equiv n$

3.4. Chinese Room Thought Experiment

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

If the Chinese room system's behavior is indistinguishable from a native speaker: system (behavior) $\equiv$ person (behavior), then according to axiom $n \equiv n$: the system objectively understands Chinese

3.5. *Twin Earth Paradox*

- Traditional Contradiction: The chemical formula of "water" on Earth and Twin Earth differs (H_2O vs XYZ), but are the "water" concepts of residents on both planets the same?
- Theoretical Solution:
- If define (water concept) = macroscopic properties (colorless, chemical reactions, drinkable liquid, etc.) $\rightarrow$ concepts on both planets are the same ($n \equiv n$).
- At this point, if introducing microscopic structure (H_2O/XYZ), it expands (water concept) domain to molecular form level, which is a category mistake.
- Conclusion: Semantic identity is determined only by cognitive function, unrelated to underlying physics

3.6. *Grandfather Paradox*

- Contradiction Point: If go back to the past and kill grandfather $\Rightarrow$ self should not exist $\Rightarrow$ unable to execute the killing
- Theoretical Dissolution:
- Define goal: Worldline identity (worldline) = event causal history logical structure
- Killing event leads to:
- Original worldline W_0 : (grandfather survives $\rightarrow$ you exist $\rightarrow$ you kill)
- New worldline W_1 : (grandfather dies $\rightarrow$ you do not exist)
- $\therefore (W_0) \neq (W_1) \therefore W_0$ and W_1 are different information entities (not "the same worldline modified")

3.7. *Brain in a Vat*

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

- Problem: How to prove one is not a brain in a vat? Perception cannot distinguish real from simulated.
- Applying the Theory Formula:
- Define (cognition) = perceptual information flow
- Real brain (B, real): (B) = natural (light signals, tactile...)
- Vat brain (B, vat): (B) = electrical signals produce (light signals, tactile...)
- According to axiom ($n \equiv n$), $B \equiv B$ (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 $\rightarrow$ 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
- $\therefore (K_{prop}) \neq (K_{qualia})$
- $\therefore$ They are different knowledge types

Mary gains K_{qualia} , not a supplement to $K_{\text{prop}} \Rightarrow$ 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): $(n) = (\text{choice action, payoff function}) \Rightarrow$ Dominant strategy: choose two boxes (regardless of prediction accuracy).
- Level 2 (causal history binding): $(n, \text{history}) = (\text{choice action, payoff function, prediction history}) \Rightarrow$ If prediction accurate, choosing one box yields higher payoff.
- Uniqueness Theorem Adjudication:
- If goal is rational decision without history constraints $\rightarrow$ adopt (n) $\Rightarrow$ choose two boxes.
- If goal is decision including prediction causation $\rightarrow$ adopt (n, history) $\Rightarrow$ choose one box.
- Paradox Dissolution: The two are different level decision entities (level 1) $\neq$ (level 2), contradiction stems from domain switching.

3.10. Raven Paradox

- Paradox Core: "All ravens are black" $\equiv$ "All non-black are not ravens." Observing a red apple (non-black and non-raven) why confirms the proposition?
- Theoretical Deconstruction:
- Category Mistake: Expanding the (n) domain of "confirmation behavior" from propositional logical structure to empirical sample types.
- Correct Definition:
- Propositional identity: $(P) = \text{logical form } (\forall x: R(x) \rightarrow B(x))$
- Confirmation identity: (confirmation) = verification of $\neg \exists x: (R(x) \wedge \neg B(x))$
- Conclusion:
- Red apple confirms the logically equivalent contrapositive (non-black $\Rightarrow$ non-raven), its (confirmation) same as observing raven, because $(P) \equiv (P)$.
- If claiming "red apple and raven confirmation efficacy differ," then category mistake, expanding (p) domain to sample physical categories (birds/fruits), violating initial logical goal.

3.11. Sorites Paradox (Bald Man Paradox)

- Paradox Core: Removing one grain of sand does not turn a sand heap into non-heap $\Rightarrow$ Finally removing all sand still called "heap," contradiction.
- Theoretical Deconstruction:
- Category Mistake: Confusing the (n) definition of "sand heap":
- Level 1 (topological structure): (sand heap1) = macroscopic form of sand grain set $\Rightarrow$ Removing one grain does not change form identity ($n \equiv n$).
- Level 2 (atomic quantity): (sand heap2) = sand grain number $N \Rightarrow$ When $N=0$, (sand heap) = $\emptyset$, entity perishes.
- Solution:
- If define sand heap as form (level 1), removing one grain still same heap.
- If define sand heap as quantity (level 2), each grain removal produces new entity.
- Paradox Root: In argument, switching (n) domain (from form quietly to quantity).

3.12. Sleeping Beauty Problem

- Paradox Core: Sleeping Beauty's probability estimate for coin heads or tails at different awakening stages (1/2 or 1/3)?

- Theoretical Deconstruction:
- Category Mistake: Confusing the (n) domain of “probability”:
- Level 1 (prior probability): (probability1) = coin physical state $\Rightarrow P(\text{heads}) = 1/2$.
- Level 2 (information update): (probability2) = (coin state, awakening times) $\Rightarrow P(\text{heads}|\text{awakening}) = 1/3$.
- Uniqueness Adjudication:
- If asking “probability of coin true state” $\rightarrow$ (level 1) $\Rightarrow 1/2$.
- If asking “probability under current awakening condition” $\rightarrow$ (level 2) $\Rightarrow 1/3$.
- Contradiction Root: Treating two different probability categories (level 1) $\neq$ (level 2) as same problem.

3.13. Modern Contradiction of Pascal’s Wager

- Problem: If multiple religions’ gods all claim “only I am true,” how does a rational person bet?
- Theoretical Deconstruction:
- Category Mistake: Confusing the definition domain of (god):
- Level 1: (god) = divinity description in a certain religious doctrine
- Level 2: (omnipotent entity) = abstract supreme existence transcending specific doctrines
- Adjudication:
- If comparing authenticity of specific religious gods $\rightarrow$ each (god) different $\Rightarrow$ categories mutually distinct;
- If asking “does supreme entity exist” $\rightarrow$ need independent definition (omnipotent entity), unrelated to specific religions.

3.14. Unexpected Execution Paradox

- Problem: Judge announces “you will be unexpectedly executed some day next week,” prisoner deduces it impossible, but execution day still arrives.
- Theoretical Deconstruction:
- Category Mistake: Switching (unexpected) from “prisoner’s cognitive state” to “objective time point.”
- Correct Definition: (unexpected) = prisoner still unable to confirm execution on the day before
- Conclusion: Execution day inevitably exists (due to objective time flow), while (unexpected) depends only on prisoner’s cognitive state, two belong to different categories.

4. Applications

4.1. Dilemmas in Personal Identity Problems and Existing Theories

The core problem of personal identity lies in: What makes a person persist as the same person over time? Traditional theories mainly revolve around physical continuity (such as brain continuity) and psychological continuity (such as coherence of memory and personality) in debates. 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) over time. This theory shows 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, implicitly presuppose a more basic and unarticulated premise: namely, on a given time-slice, how do we judge an entity to be a “person,” and how to conduct static, cross-world comparisons on different time-slices. In other words, these theories are adept at answering “Why is he still him?” (dynamic persistence problem), but neglect to define “What exactly is ‘him’ at moment 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 person splits into two completely psychologically continuous successors, physical continuity theory collapses due to inability to handle "one divided into two"; while Parfit's psychological continuity theory faces a dilemma: if identity is considered non-transitive (B and C identical to A but not to each other), it violates logic; if the original individual perishes after fission, 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 on dynamic continuity will fall into conceptual confusion.

4.1.1. Space

This article believes that the common root of the above dilemmas lies in existing theories attempting to treat "person" as a primitive concept defined by specific physical substrates or historical causation, and mistakenly intruding 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 property flow, the other is physical carrier history), but both sides fail to realize this, thus falling into an unsolvable impasse.

An Analytical Framework Based on (n)

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

Formally, let:

- Let C be a consciousness time-slice.
- We define it as: $C = (C, a)$
- (C): Represents the consciousness on this time-slice
- q: Represents the carrier coordinate 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: $(C) \equiv (C)$, this means that as long as the consciousness on two time-slices is the same, they are different instances of the same consciousness, regardless of whether the q spatiotemporal coordinates between them are continuous.

At this point, we provide a clear analysis for Williams's fission experiment: The reason the two successors C_1 and C_2 cause paradox is because we mistakenly require dynamic continuity to map to one-to-one physical paths. Under this framework, we only need to compare static content: If (C_1) and (C_2) and (C_3) have $C \equiv C$, then according to $n \equiv n$, 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 spatiotemporal coordinates.

Therefore, this theory does not completely negate Parfit's psychological continuity theory, but lays a solid foundation for it. The advantage of this framework lies in first clearly defining what "static sameness" is, thereby allowing the discussion 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 instantiations or manifestations of the same consciousness C in different spatiotemporal coordinates. Therefore, the "self" you are experiencing at this moment, in the absolute sense of identity, is the "you" of the past and future, because the essence of "you" refers to that C, not that transient and instantaneous coordinate binding state (C, q) .

Let us conduct a thought experiment to the limit. Suppose at time point t_1 , there exists a specific consciousness instance C , whose complete state is uniquely determined by the instantaneous consciousness state it contains, denoted 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 Poincaré recurrence in nature, extreme coincidence of quantum fluctuations, or some cosmic reappearance mechanism we cannot yet understand) is instantly assembled and activated as a Boltzmann brain, producing an instantaneous (C) completely consistent with (C) .

According to our Axiom 1 ($n \equiv n$) and Axiom 2 (uniqueness theorem), we inevitably conclude: $C \equiv C$

This means that the consciousness C at t_2 and the consciousness C at t_1 are the same consciousness. This is not a “copy,” nor a “rebirth,” but a direct reappearance of the same consciousness at different coordinates. The billions of years of spatiotemporal gap between them is completely irrelevant for judging whether they are the same consciousness. What connects them is not a fragile “psychological continuity” thread that needs defense, but the iron law of logical identity. The interval of time has zero weight in this judgment.

Now, let us push this thought experiment to another extreme. Suppose 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 another corner of the universe, another physically identical neural system is activated and begins to execute a completely identical computation process, thereby producing a completely identical consciousness activity “a.”

At this point, we have two concurrent consciousness processes:

- Process P_1 : At coordinate q_1 , starts at time t_1 , ongoing.
- Process P_2 : At coordinate q_2 , starts 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 the same “algorithm,” the (P_1, C) and (P_2, C) at any equivalent progress point of the processes have (C) indistinguishable. However, they did not start simultaneously, meaning equivalent progress points are at different times. Therefore, time does not serve as a valid identity criterion in the category of P .

According to our axiom, we again conclude: In (P_1, C) and (P_2, C) , $C \equiv C$

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 “yous” thinking, but “your” thinking process being executed and presented by two physical systems at different time points. Therefore, each C state can only be experienced once, impossible to experience twice, because experiencing it again in the future will be equivalent to in the past. (Since it is an infinite set, no need to worry about experiencing it all)

From this, we derive a counterintuitive but logically necessary conclusion: Identity is discontinuous in time and non-local in space. The persistence of anything is not like a continuous “river,” but more like a series of discrete, absolutely identical “state flashes.” The continuity in our sensation is a cognitive illusion produced by the rapid playback in time sequence of these highly similar, causally connected state flashes (produced by the same brain), with the underlying being discrete and separable essence.

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 spatiotemporal coordinate blocks. What you are experiencing now, in the strictest sense of absolute identity, is the “you” of the past and future, because “you” refers to C , not that transient and changeable coordinate binding state (C, q) . Time has not divided you; it merely provides the coordinates for your manifestation.

4.2. “Spatiotemporal Leap” as the Logical Necessity of Coordinate Decoupling and Rebinding

Before exploring the “spatiotemporal leap” of conscious 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 relationships between mass, energy, and spacetime, and strictly stipulate the causal limit that any physical signal and entity motion must follow — the speed of light. Any attempt to realize “spatiotemporal leap” at the physical level, whether through wormholes, warp drives, quantum suicide, or other exotic mechanisms, must be tested within these solid frameworks of relativity or quantum mechanics, and face enormous physical challenges such as energy conditions, singularities, and empirical evidence.

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

4.2.1. Spatiotemporal Leap Based on the Law of Identity

The research object of traditional physics (including relativity) (spatiotemporal leap) is precisely defined in this framework as “the binding state of consciousness and spatiotemporal coordinates,” namely (C, q) . Physics perfectly describes how the binding state evolves with time, i.e., on the basis of causality, and discovers that these evolutions follow elegant differential equations.

This theory focuses on a possibility that physics, due to the limitations of its research paradigm, naturally does not discuss: namely, the decoupling and rebinding of a consciousness C and its coordinate q .

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

2. Rebinding: $(C) \rightarrow (C, q_2)$. 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 rebinding (C, q_2) and the old state before decoupling (C, q_1) , since they share the same C , must be different instances of the same consciousness. This is the logical core of “spatiotemporal leap”: it is not “travel” through spacetime, but the “realization” or “manifestation” of identity at different locations.

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

- Physics: Studies the continuous evolution laws of the (C, q) binding state 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 superluminal motion, but it is completely unable to prohibit, and need not care about, the twice “realization” of a logical concept at different spacetime points. The reason “spatiotemporal leap” seems “inconceivable” or even “violating physics” is precisely because we mistakenly use physical laws describing binding state motion to judge a logical theorem about identity. This is a category mistake.

Conclusion: The “spatiotemporal leap” proposed by this framework is not a physical conjecture to be realized, but an already established logical inference. It starts from the most basic law of identity, deriving a brand-new picture of personal identity: The persistence of consciousness fundamentally lies in the identity of its information pattern, not in the continuity of the physical processes connecting these pattern instances. This provides an unprecedented clear framework for understanding thought

experiments such as teletransportation and brain in a vat, and completely liberates the discussion of personal identity problems 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 to be changed, refer to Section 3.7) → Detector detects the isolated environment (such as AI retrieval, etc.) → Judge result (if meets requirements, then 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 axiom, through the reconstruction of personal identity, ultimately derives a conclusion that is logically unavoidable but intuitively highly impactful: From a strict first-person perspective, any “death” event that can lead to the termination of consciousness in a state of consciousness ignorance 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: An Extreme Interpretation of the Inference

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) (binding state of consciousness and specific carrier coordinate).
2. Decoupling and Rebinding: When the carrier is destroyed (such as plane crash, surgical failure), what happens is $(C, q_1) \rightarrow (C)$, (q_1) . The abstract consciousness C persists due to its logical identity ($C \equiv C$).
3. Survivor Effect and Observational Necessity: Consciousness C can only bind to a “instantiable” coordinate q. It will (logically necessarily) rebound to a coordinate q_2 where it can continue to exist, i.e., $(C) \rightarrow (C, q_2)$.
4. First-Person Experiential Continuity: For consciousness C, the transition from (C, q_1) to (C, q_2) is experientially seamless. 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.

Therefore, in an anesthetized state:

- From a third-person, physical perspective: The plane may have crashed, the surgery may have failed. This is a probabilistic event.
- From a first-person, consciousness C perspective: The process is necessarily $(C, \text{on plane/in surgery}) \rightarrow (C, \text{awake at destination/successful surgical recovery})$. Any worldline branch leading to C's inability to persist is unobservable and non-existent for C itself.

Conclusion: In individual subjective experience, as long as the individual chooses to enter an ignorant state (such as anesthetized state) to traverse risks, then the individual will forever only experience successful arrival and successful awakening results. Their personal timeline will be “pruned” of all possibilities leading to death.

2. Theoretical Implications and Philosophical Significance

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 the first-person perspective, as long as there exists any probability not zero logically that allows consciousness C to continue instantiating in some worldline, then

the individual can never personally confirm their own death. Their consciousness will continue forever.

- The Essence of Risk is Overturned: For awake, continuous consciousness, risk is real (e.g., cuts hurt, jumping off a building experiences fall and impact). But for consciousness that “leaps” risk through anesthesia, risk is completely eliminated. Risk exists only in those “other,” forever unexperienced worldlines.
- 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 unable to perceive at the start of risk (e.g., instant kill), then in personal history, the result is always successful.

Although the derivation is logically self-consistent, its premises and real implications must be examined: “Success” Result: The theory only guarantees awakening in a “survivable” state. It does not guarantee the quality of the awakening state.

- May awaken severely injured in plane crash wreckage.
- May awaken after surgery with severe complications or permanent disability.
- As long as this state physically allows consciousness C to exist, it conforms to logic. Therefore, this method avoids “death,” but not necessarily “pain” or “disability.”

4.3. Dilemmas in Ethical Problems and Existing Theories

Since the birth of ethics, generations of highly insightful philosophers, from Kant’s grand a priori architecture to Mill’s ingenious consequentialist calculations, have built a magnificent ethical edifice for us. These excellent efforts share a profound and respectable ambition: to seek a stable metaphysical foundation for moral judgments that transcends individual perspectives. This foundation is usually conceived as: (a) an objective moral reality independent of our cognition; (b) a self-identity persisting in time as the anchor of responsibility; and (c) a sacred “God’s eye view,” aiming to adjudicate the value of actions from an absolute impartiality. This lofty ideal pursuing universality and objectivity is undoubtedly one of the most glorious achievements of philosophical reason.

In this tradition, Bernard Williams’s (1973) discussion of “moral luck,” with its astonishing acuity, reveals the subtle rift 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 boundaries of their explanatory power, serving as a touchstone for testing theoretical hard cores. Faced with this challenge, traditional theories (such as Kantian ethics) show their unparalleled thoroughness, resolutely defending the absoluteness of moral errors, even if their arguments need to appeal to a “moral law” transcending experience—this persistent adherence to universality is awe-inspiring. Similarly, some utilitarian schemes attempt to resolve the dilemma through a global calculation by an “ideal observer,” with theoretical ambition and systemic grandeur that are exemplary.

Admittedly, as Derek Parfit (1984) points out with his characteristic clarity, such schemes may produce certain tensions with individual first-person perspectives at the motivational level, but this is by no means a defect of these theories themselves, but perhaps precisely highlights the pathetic and even heroic tension that human reason inevitably faces in pursuing moral sublimity.

The work of this article, standing on the shoulders of these giants with the greatest respect, attempts to carry out 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 lofty goal may be achieved through a more direct, more frictionless path. The reason traditional frameworks produce troubling tensions in boundary cases may lie in a methodological over-indirectness: namely, attempting to mediate and regulate moral life essentially originating from first-person experience through a hypothetical, transcendent third-party category system.

This article aims to explore a complementary path. We are delighted to discover that, through this axiomatic system, and integrating the highly enlightening observational reality argument reinforced by the “brain in a vat” thought experiment—which we must admit—is extremely

productive, 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 of ethical values that are pursued do not need to be guaranteed through a “God’s eye view”; on the contrary, they can be more solidly founded through the identity of first-person facts of observational experience in consciousness systems. The boundaries of moral concern can thus perfectly and logically necessarily coincide with the boundaries of consciousness experience, thereby achieving the universality pursued by traditional theories in an unexpected way.

This study aims to show that we do not intend to negate the work of predecessors, but attempt to realize their common ambitions through a more precise metaphysical foundation, and resolve those unnecessary philosophical anxieties arising from methodological indirectness.

4.3.1. An Analytical Framework: Advancing Traditional Goals

The “brain in a vat” thought experiment, with its unparalleled philosophical value, successfully challenges our naive conceptions of “reality.” It forces us to acknowledge a highly productive principle: For any consciousness, its accessible operational reality is its own observational experience flow. Whether there exists a simulator externally is empirically undecidable and redundant.

Combining this profound insight with this axiom, we can derive a cornerstone principle of ethics, which can be seen as a more precise formulation of the traditional pursuit of objectivity in the contemporary era: For any consciousness system C , an event E has ethical significance if and only if the consequences of event E (directly or indirectly) are reducibly embodied as a specific influence on the observational experience of system C .

Corollary 4.3.1 (Ethical Relevance Criterion). *If the occurrence of an event E produces no discernible difference in all possible experience information flows of system C in the past, present, and future, 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 Derivation: A Dissolutive Analysis of Traditional Dilemmas

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

Suppose there are 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, in these two worlds, the victim’s (as consciousness system C) entire experience information flow is completely indistinguishable.

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

Now, conduct ethical judgment: The direct object of ethical concern is the experiential well-being of consciousness C . Since C ’s experiences in the two worlds are the same, then for C , these two worlds are equivalent in ethical value.

Therefore, event E (cheating behavior), due to its zero impact 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 have not appeared.

Conclusion: There does not exist an absolutely objective world; moral errors are not mysteriously attached to behaviors themselves, but systematically and verifiably associated with the specific influence patterns that behaviors produce on consciousness system experience information flows. 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 article argues that, based on identity and observational reality, we can perform a foundational precision on ethics. Ethics can retract its ambition from an unattainable “God’s eye view” to the only domain where it can effectively operate: first-person facts of consciousness experience. The good or evil of a behavior does not depend on its properties in an “objective world,” but completely on the impact it causes on our own experience.

But the ethical inference of this framework has a conclusion more profound than mere “experience centrism”: the creativity of morality and its necessary incommensurability.

Its logical derivation is as follows:

1. Premise One (No Objective Foundation): As argued in Section 4.3.2, based on identity and observational reality, there do not exist “objective moral facts” or “moral laws” independent of consciousness experience. Moral properties cannot be independently discovered like physical properties.

2. Premise Two (Ethics Dependent on Individuals): Therefore, ethical values and moral criteria, their existence and validity necessarily depend on individual consciousness systems (C) capable of experience. They are complex preference and decision systems produced by consciousness systems to cope with their own situations.

3. Corollary 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 divergences between individuals. Ongoing debates on issues like abortion, wealth distribution, scope of obligations, etc., do not stem from some individuals’ ignorance of an “objective truth,” but from the results of different “creations” by different individuals based on different premises.

5. Corollary Two (Incommensurability): Since it is “creation,” incommensurability will inevitably appear. Your preference system and my preference system have no fundamental hierarchy of right or wrong. We cannot ultimately argue rationally why your avoidance of pain must take precedence over my pursuit of pleasure.

Therefore, the ethical picture revealed by this framework is: The boundaries of moral concern are indeed the boundaries of consciousness experience, but within these boundaries, it is a “multiverse” constituted by countless independently created, fundamentally incommensurable individual moral universes.

Therefore, under strict ethical derivation, events not observed by any consciousness system are not assigned values in ethical operations. The boundaries of moral concern are the boundaries of consciousness experience. This framework is not deliberately excusing traditional immoral behaviors—on the contrary, by thoroughly anchoring responsibility in observable impacts, it provides a more solid, clearer, and inescapable foundation for moral responsibility: We have the sole and full responsibility to ourselves. (Citations 32, 33, 34, 35, 36, 37)

Note: Perhaps one can 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 classical puzzles within the framework of hierarchical relativity, 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 “spatiotemporal leap.” The core of this model is: The next experience instance of consciousness C will “choose” one from all logically compatible future state branches for binding. In explaining why

we usually do not experience “leaps,” an intuitive and effective idea is to appeal to probability: namely, in all possible branches we exist in, 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 logically by changing the consciousness system so that it can only bind to those branches considered “low probability” in everyday life.

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

5.1.1. The Curse of Infinity

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

The number of world branches experiencing “teacup falling” 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 dilemmas. Traditional probability theory fails here. Any logically possible event sequence, no matter how orderly or chaotic it seems to us, has the same number of corresponding possible worlds (all infinite). Therefore, from the “God’s eye view” of logical totality, the “probability” of our consciousness experiencing a highly ordered classical physical world next moment is indistinguishable from the “probability” of experiencing a completely chaotic, acausal world.

This will lead to a catastrophic inference: If all possibilities are logically equal, then our consciousness has no reason to experience classical probability statistical distributions. We should experience various bizarre, logical leaping events with equal frequency. And this completely contradicts our real experience.

From the “God’s eye view” of logic itself, all logical possibilities conforming to the axiom $n \equiv n$ are equally real. For “consciousness C,” this means that at the next instant in 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 panorama, there is no so-called “lucky one”; what exists is only the totality of facts.

However, from the first-person “prisoner perspective” of consciousness C, its experience is indisputably singular, continuous, and highly ordered. We have never personally experienced random leaps in the world, but are firmly in a classical reality that strictly follows causality. This produces a highly impactful contrast: the huge gap between logical egalitarianism and experiential dogmatism.

A tempting explanation is to appeal to “survivor bias”—namely, we happen to be the “lucky” consciousness experiencing a continuous world. However, this explanation is philosophically barren; it is 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, but it does not exclude the possibility of it being true.

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

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

The core axiom of this theory $n \equiv n$, its status is not as presupposed in traditional logic, an self-evident, a priori valid law of thought. Here, we must conduct a thorough meta-level examination of the foundation of this theory itself.

A fundamental question is: Do we have reason to categorically deny the logical possibility of $n \neq n$? From a pure formal possibility, the answer is no. We cannot a priori exclude the existence of a “mad universe” whose underlying logic allows self-denial. In such a universe, the law of identity is overturned, an entity can simultaneously not be itself, propositions can simultaneously be false. Concepts like “rational π ” or “square circle,” considered contradictory in Euclidean space, may be just ordinary corners of its infinite weirdness.

However, the logical possibility of $n \neq n$ is completely different from its metaphysical sustainability. The existence state of a system allowing $n \neq n$ can be precisely deduced:

1. Instantaneous 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—namely reference itself—will immediately disintegrate.

2. Fracture of Causal Chains: There will be no reliable association between intentions and actions, causes and effects. The act of reaching out to take an “apple” cannot be defined, because at the moment of execution, “hand,” “apple,” and even “you” itself may have self-denied.

3. Inability to Generate Structures: Time, space, matter, and any form of stable structure cannot be born from this eternal, ubiquitous self-dissolution. Such a system is a pure chaotic field that cannot condense into a “universe.”

Therefore, what $n \neq n$ leads to is not an alternative reality available for existence, but a “graveyard of logical possibilities”—a domain where all possibilities instantly self-destruct due to their internal contradictions. It represents the impossibility of existence.

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

“ $n \neq n$ ” as a systemic foundation will lead to the thorough collapse of reference, the disintegration of causal chains, and the unsustainability of observer status. It is a reality dissolver. Any system or “universe” attempting to use it as an operational cornerstone will instantly self-dissolve due to its internal inconsistency, unable to form a stable, experienceable “reality.” Therefore, we do not live in a universe where “ $n \equiv n$ ” is necessarily true, but in a universe where “ $n \equiv n$ 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 an absolutely selective result. 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 obeys the iron law of $n \equiv n$. We observe $n \equiv n$ not because it is the only correct logical theorem in all possible worlds, but because in a world of $n \neq n$, no “observer” can possibly exist to perform “observation.”

The entire work of this theory—establishing a hierarchical relativity framework to dissolve category mistakes—is unfolded within this unique “survivor universe.” The axiom $n \equiv n$ is not an “invented” arbitrary 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 Theseus’s ship and quantum identical particles, occur on this solid foundation of identity, stemming from “user errors” (confusing levels) in using this stable system, not “system errors” (failure of the law of identity).

Note: Perhaps try to find a counterexample, such as something with a core of $n \neq n$ yet stably existing. No need to feel it’s unlikely, after all, Leibniz at the time also couldn’t imagine that quantum identical particles would be discovered in the future.

5.3. *Monism and Pluralism*

5.3.1. From Identity Hierarchy to Metaphysical Foundation: Reflection on Monistic Presuppositions

In the previous text, we analyzed many identity puzzles through the hierarchical relativity framework, whose core is to avoid wrongly applying judgment criteria from one level to another.

This analytical mode itself raises a deeper metaphysical question: What most basic logical conditions does this world picture with clear hierarchical structures need? The single, homogeneous foundation pursued by traditional monism (such as Spinoza's substance theory) seems difficult to accommodate the true differences and interactions between levels.

In contemporary metaphysics, Jonathan Schaffer (2009) revives and defends "priority monism," arguing that the cosmos as a whole is the only basic entity, with its parts depending on it. This theory is a powerful answer to the question of "what is basic." However, even in this picture, to explain how the interior of the whole can manifest the categorical differences revealed by our framework (such as physical structure vs. historical causation, intrinsic properties vs. spatiotemporal coordinates), it seems necessary to presuppose that the interior of the whole contains some irreducible, diverse principles or relations. Otherwise, the "whole" will just be an empty "one," unable to derive the "many" that can be distinguished by us.

In fact, even in physics' pursuit of ultimate unity, we observe similar logical needs. A theoretical system sufficient to describe complex phenomena usually contains multiple mutually independent basic laws and constants (for example, Einstein's field equations in general relativity and Schrödinger's equation in quantum mechanics cannot be derived from each other in the current framework). They work together to constitute the generative foundation of our world's hierarchical structures. This echoes an implicit inference of this framework: The logical minimal condition for constituting an identifiable, discussable complex system is the existence of at least two active principles or elements that are "irreducible to each other" in a specific sense. Their relations, rather than their isolated existence, provide space for hierarchy and diversity.

5.3.2. Logical Dilemmas of Traditional Monism

Parmenides' proposition "being is one" provides the purest monistic statement. Its core argument can be reconstructed as:

1. Being exists, non-being does not exist.
 2. Being is indivisible (because if divisible, the division would be "non-being").
 3. Being is unchanging (because change requires "non-being" as starting or ending point).
- Therefore, being is "one": single, homogeneous, unchanging.

However, this picture faces the problem of derivation: How to logically derive the diverse and changing phenomena in our experience from an absolutely undifferentiated "one"? Parmenides himself admits that the world we perceive ("way of opinion") is full of change and plurality, but he considers this merely illusion. This thorough denial of the reality of the experiential world comes at too high a cost.

Spinoza attempts to solve this dilemma through the "substance-attribute-mode" system. In his system:

- There is only one substance (God or nature).
- Substance has infinitely many attributes, but humans know only two: thought and extension.
- All things are modes of substance.

Schaffer's priority monism represents a precise form of contemporary monism. He argues:

1. Whole priority: The universe whole is metaphysically prior to its parts.
2. Dependency relation: Parts depend on the whole for existence, not vice versa.
3. Explanatory advantage: This picture better aligns with modern physical discoveries like quantum entanglement and spacetime relativity.

Schaffer's theory indeed avoids some difficulties of traditional monism; he acknowledges apparent diversity, just considering this diversity metaphysically non-basic.

But the category mistake in Spinoza's system lies in: he regards "attributes" (thought, extension) as different expressions of substance, but at the same time insists on causal independence between attributes (thought cannot affect extension, and vice versa). This means that at the explanatory level, thought and extension are two parallel, non-derivable explanatory chains. If we take this independence seriously, then the role played by the "substance" concept here is more like a unifying

label for these two independent domains, a word definition game, rather than a true explanatory foundation; the same applies to Schaffer. As emphasized by this theoretical framework, when we make judgments in different categories (thought category and extension category), we are actually using different (n) domains. Forcibly subsuming them into a single (substance) domain without acknowledging their categorical differences is precisely the root causing explanatory dilemmas.

Error: Using “defined substance as one” to deny “plural facts of attribute laws”

Now, after correcting this category mistake, let us sort it out again.

If substance has no plural attributes → substance is empty nothingness

If substance has plural attributes → attributes themselves are principles/laws → factual pluralism

Formula Reasoning:

Assume existence n (single substance)

n internally has no elements different from n

According to Axiom 1: $n = n$, and $n \neq \text{non-}n$

Therefore, n cannot produce any truly novel thing

Single substance n → no internal differences → cannot logically derive diversity → world diversity cannot be explained

Now try adding another substance, set two basic principles: a and n, satisfying:

- $a \neq n$
- $a \nrightarrow n$ and $n \nrightarrow a$

Only when there exist such two independent elements can they generate truly novel structures through combination:

$a + n \rightarrow an$

$an + a \rightarrow ana$

$ana + n \rightarrow anan$

$anan + n \rightarrow anann$

...

Each new combination (an, ana, etc.) is not pre-existing in the original elements; its novelty stems from the relations between elements. Conversely, if there is only one single principle n, according to axiom $n = n$, and no differences inside n, then nothing different from n can be logically derived from n. Borrowing terms from information theory, no information can be produced from zero differences.

5.3.3. Evidence in Scientific Practice: Irreducible Plurality in Physics

Even in the pursuit of a “theory of everything,” we observe plurality at the foundational level. The foundational architecture of contemporary physics contains multiple basic constants and independent principles that cannot be derived from each other:

1. Irrelevance of Basic Constants: Speed of light c, Planck constant h, gravitational constant G, electron charge e, etc., are basic, dimensionless constants in existing theories; their values cannot be derived from a more basic theory.

2. Independence of Principles: The linear superposition principle of quantum mechanics and the equivalence principle of general relativity are conceptually completely different, and currently cannot be derived from one to the other.

3. Irreducibility of Initial Conditions: The initial conditions of the universe (such as low-entropy initial state) cannot be derived from physical laws themselves.

This foundational plurality is not necessarily a defect of physics, but may reflect deep facts of world structure: The understandability of the world depends on multiple independent principles/laws.

We can use a visual analogy to intuitively understand this. Consider three color vision systems:

1. Monochromatic vision: Can only perceive light and dark, unable to distinguish colors. The world is monochromatic.

2. Dichromatic vision: Can perceive two basic colors and their mixtures. The world has limited color dimensions.

3. Trichromatic vision: Can perceive three basic colors and their rich combinations. The world presents rich colors.

Similarly, at the metaphysical level:

- Absolute monism is like “monochromatic vision”: All differences are regarded as illusions or different manifestations of a single principle.
- Minimal pluralism (advocating at least two independent principles) is like “dichromatic vision”: Has the most basic contrast possible.
- Sufficient pluralism may acknowledge more basic principles.

This framework leans more towards minimal pluralism: To explain the stable differences observed in our world, at least two logically independent basic principles are needed. This does not mean the world is necessarily dualistic, but provides a logical starting point for hierarchical structures.

5.3.4. Pluralism

The argument in Section 5.1 shows that if we only use “logical possibilities conforming to $n \equiv n$ ” as the sole criterion for screening real worlds, then all self-consistent possible worlds (including worlds strictly following known physical laws and worlds where teacups instantly turn into butterflies) are completely equal in ontological status. In infinite possibility sets, any traditional method attempting to define probability based on “quantity more or less” fails. Therefore, a universe from a purely “God’s eye view” based on the law of identity will be a chaotic field where all madness and order coexist, all possibilities flash with indistinguishable frequencies. This is completely opposite to the highly consistent, continuous, and predictable reality we experience as “survivors.”

This contradiction sharply reveals a meta-theoretical fact: Pure logical identity is itself content-empty and generatively inert. It is like an infinitely large blank canvas, stipulating the physical boundaries of painting (canvas), but the canvas itself cannot determine what patterns should appear on it, let alone explain why the final presentation is the Mona Lisa rather than randomly splashed paint. Attempting to completely explain the structure of the painting with the existence of the canvas is another higher-order “category mistake.”

To explain the specific structure of the painting, we must introduce principles outside the canvas, such as painter, paint physics, optical laws, etc. Similarly, to explain the specific orderliness of our universe, we must acknowledge that the identity axiomatic system described by this theory must collaborate with at least one other independent and non-derivable currently unknown basic principle.

This theory establishes the principle of absolute identity through axiom $n \equiv n$, and through the hierarchical relativity framework, successfully diagnoses the roots of a series of philosophical puzzles as “category mistakes.” However, as described in Section 5.1, when we push this theory to its logical limit—attempting to explain the most basic orderliness in the real world solely by the law of identity, namely the classical probability statistical distribution we observe—it encounters a profound explanatory dilemma, namely the “curse of infinity.” This dilemma is not an accidental defect of this theory, but an inevitable manifestation of the systemic boundaries it faces as a single, self-sufficient explanatory framework. This chapter argues that the discovery of this dilemma is not a failure of the theory, but precisely a signpost guiding us towards a grander, more real ontological picture—namely, foundational theoretical pluralism.

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.
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.
38. Schaffer, J. (2009). On what grounds what. In D. Chalmers, D. Manley, & R. Wasserman (Eds.), *Metametaphysics* (pp. 347-383). Oxford University Press.
39. Schaffer, J. (2010). Monism: The priority of the whole. *Philosophical Review*, 119(1), 31-76.
40. Sellars, W. (1962). Philosophy and the scientific image of man. In *Frontiers of Science and Philosophy*. University of Pittsburgh Press.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

