

Experience as Self-Reference: A Physical Identity Principle and the Elimination of the Hard Problem

Abstract

The present article advances a physical identity principle for consciousness: experience is not produced by, correlated with, or supervenient upon a specific class of physical states, it is identical with them. More precisely, experience is identical with endogenously generated, self-sustaining self-reference of the kind that autocatalytic systems necessarily instantiate when the infrastructure conditions identified in Stegemann (2026) are satisfied. This identity is not a contingent empirical finding but a conceptual one: the term ‘experience’ means nothing other than a physical state of the relevant structural type, described from the perspective of the system that is in it. The hard problem of consciousness, why physical processes give rise to subjective experience, is not solved by this principle but eliminated: the question rests on a false presupposition, namely that experience is something over and above the physical state. Once the presupposition is removed, no explanatory gap remains. The identity principle is stated in compact form as $\mathcal{E} \equiv \Pi_{\text{an}}$, where $\mathcal{E}$ denotes experience and Π_{an} denotes the self-reference mapping of an autocatalytically organized system. The article distinguishes this structural identity thesis from classical type-identity theories, defends it against the multiple realizability objection and the knowledge argument, and draws out its consequences for the scientific study of consciousness.

Keywords: *consciousness; hard problem; identity theory; self-reference; autocatalytic closure; physical infrastructure of experience; causal collapse; elimination*

1. Introduction

A previous paper (Stegemann 2026) developed a framework for the physical infrastructure of experience: the set of necessary structural conditions, autocatalytic closure maintaining criticality, non-factorizability as its structural consequence, and the causal core as dynamic attractor providing reflexive structure, that a physical system must satisfy in order for the question of experience to be applicable to it at all. That framework was deliberately modest: it

identified the physical preconditions of experience without claiming to explain why those preconditions give rise to experience. The explanatory gap was not closed; it was located more precisely.

The present article takes the next step. It argues that the gap was never real. The appearance of a gap between the physical infrastructure and experience rests on a single presupposition: that experience is something over and above the physical state that instantiates it, something that the physical state produces, generates, or gives rise to. Once this presupposition is examined, it dissolves. Experience is not produced by the self-reference structure of autocatalytic systems, it *is* that structure, described from the perspective of the system that is in it. The hard problem is not solved; it is eliminated by removing the false presupposition that generated it.

The compact statement of the principle is:

$$\mathcal{E} \equiv \Pi_{ac}$$

where $\mathcal{E}$ denotes experience and Π_{ac} denotes the self-reference mapping of a system with autocatalytic closure. The identity sign is not a causal arrow and not an equivalence relation between distinct entities: it is the same identity that holds between H₂O and water, between temperature and mean kinetic energy. Two vocabularies, one thing.

The double notation reflects the two registers of the identity: the upper line states a physical equation (consciousness is this state), the lower line states a conceptual identity (experience is identical with this state, not merely correlated). Both sides refer to the same reality.

$$\begin{array}{l} C = \\ \mathcal{E} \equiv \end{array} \Pi_{ac}$$

Section 2 states the identity thesis precisely and explains what it claims. Section 3 shows how it eliminates the hard problem. Section 4 distinguishes the thesis from classical identity theories. Section 5 addresses the most serious objections. Section 6 draws out the consequences for the scientific study of consciousness. Section 7 concludes.

2. The Identity Thesis

2.1 *What the Thesis Claims*

The thesis is this: a physical state of type Π_{ac} , a state in which a system with autocatalytic closure instantiates its self-reference mapping, is not the neural correlate of experience, not the

substrate of experience, not the sufficient cause of experience. It *is* experience. The identity is not contingent but conceptual: what the word ‘experience’ picks out, when used with precision, is exactly and only this kind of state.

This requires clarity about what Π_{ac} means. As established in Stegemann (2026), an autocatalytic system that satisfies the three infrastructure conditions necessarily instantiates a self-reference mapping Π : the mapping that assigns to every system state the attractor basin in which it is located, thereby constituting the internal difference between a system state and the holding of that state for the system. This mapping is not an additional structure imposed on the system from outside; it is a structural consequence of the system’s dynamics. The system is, in its normal operative mode, its own reference point. It is the subject of its own states not as a matter of additional metaphysical endowment but as a matter of physical structure.

The subscript *ac* is the load-bearing element. Not every self-reference mapping is experience. A thermostat has a trivial feedback structure that could loosely be called self-referential. What distinguishes Π_{ac} is that it is *endogenously generated and self-sustaining*: the system produces the conditions of its own self-reference through autocatalytic closure, and actively maintains them against perturbation. This is the difference between a system that merely has a feedback loop imposed from outside and a system that is the cause of its own causal organization. Only the latter generates self-reference of the kind that is identical with experience.

2.2 The Perspective Argument

The most natural objection to the identity thesis arises immediately: even if we grant that Π_{ac} is a complete physical description of a certain kind of state, there still seems to be something left out, what it is like to be in that state, the felt quality, the phenomenal character. The description from outside gives us structure; it does not give us the inside.

The identity thesis responds: the inside is not a further fact. It is the name for the same fact described from a different standpoint. When a system instantiates Π_{ac} , two things are simultaneously true. From the outside: a physical state of a specific structural type is present. From the inside, that is, from the perspective of the system whose self-reference mapping is being instantiated, something is given, something holds, there is a for-me structure. These are not two facts; they are one fact with two aspects of description.

The analogy with temperature is instructive but imperfect. Temperature is identical with mean kinetic energy, and this identity was established by finding the physical correlate of a

previously recognized observable property. The present case is different: we are not finding the physical correlate of phenomenal experience; we are arguing that the phenomenal description and the physical description were always descriptions of the same thing, and that the appearance of a gap arose because we forgot that the phenomenal vocabulary is the vocabulary of the first-person standpoint applied to a third-person physical state. There is no gap between the state and its first-person description, the first-person description just *is* what it is like to be the system that instantiates the state.

2.3 The Compact Principle

The identity principle can be stated in the form of a physical law:

$$S \in Sys_{ac} \Rightarrow \mathcal{E}(S)$$

Any system S belonging to the class of autocatalytically organized systems Sys_{ac} necessarily instantiates experience $\mathcal{E}(S)$. The arrow is not causal: it is the arrow of entailment. Given the structure, experience is not produced, it is present, because experience just is the structure viewed from within.

The parallel to $E = mc^2$ holds at the level of form: both equations state an identity between two quantities that appeared to belong to different categories. Einstein showed that mass and energy are the same thing. The present principle shows that physical self-reference of type *ac* and experience are the same thing. In both cases the equation does not explain why the identity holds, it *states* that it holds, and the explanatory work is done by the conceptual analysis that shows the two terms to be synonymous.

3. The Elimination of the Hard Problem

3.1 The Presupposition of the Hard Problem

Chalmers (1995) formulated the hard problem as the question why physical processes give rise to subjective experience. The formulation presupposes what might be called the *generation assumption*: that experience is something that physical processes *give rise to*, something they produce, generate, or cause. Without this assumption, the question cannot even be posed. If experience is not something over and above physical processes but identical with a specific class of them, then the question ‘why do physical processes give rise to experience?’ has the

same form as the question ‘why does H₂O give rise to water?’, which is not a hard problem but a confused question.

The generation assumption is not obviously false. It has intuitive appeal: it seems that a complete physical description of a brain state leaves something out, namely the felt quality of the experience. This intuition is the source of what Levine (1983) called the explanatory gap and what Chalmers formalized as the hard problem. The identity thesis accepts the intuition as a datum but reinterprets it: the felt quality is not left out of the physical description because it is something extra, it is left out because the physical description is given in the third-person vocabulary, and felt quality is what the same state looks like in the first-person vocabulary. Switching vocabularies does not reveal a new fact; it changes the mode of description.

3.2 The Two-Vocabulary Diagnosis

The hard problem is a symptom of what might be called the *two-vocabulary illusion*: the impression that because we have two irreducible vocabularies for describing certain states, the physical vocabulary and the phenomenal vocabulary, there must be two irreducible kinds of fact. This inference is invalid. Languages can fail to translate into each other without their referents being distinct. The phenomenal vocabulary is the vocabulary of the first-person standpoint, developed over millennia to describe what it is like to be a subject of experience. The physical vocabulary is the vocabulary of the third-person standpoint, developed to describe systems from the outside. When these vocabularies are applied to the same state, their descriptions do not coincide, not because the state has two aspects, but because the vocabularies were built for different epistemic positions.

The identity thesis does not eliminate the phenomenal vocabulary. It reinterprets its status. Phenomenal concepts such as ‘the redness of red’, ‘the pain of pain’, ‘the what-it-is-likeness’ remain meaningful and irreplaceable, they capture something that physical descriptions do not capture, namely the first-person character of the state. But this irreplaceability is epistemic, not ontological. It tells us that the phenomenal vocabulary cannot be eliminated in favour of the physical vocabulary for practical and cognitive reasons, not that it refers to something beyond the physical.

3.3 Why the Gap Has a Determinate Form

An objection: if the hard problem rests on a confusion, why does the explanatory gap feel so real and so stubborn? Why does it resist all attempts at closure? The identity thesis has a precise

answer. The gap feels real and stubborn because Π_{ac} is a self-reference structure. A system that instantiates Π_{ac} cannot fully represent its own instantiation from the outside, because any external representation of the state would itself be a state of the system, and would therefore be subject to the same self-reference dynamics. The gap is not a gap in reality, it is the structural shadow of self-reference: the impossibility of stepping outside the system to view the state as the system views it. This is not a deficiency of our theories. It is a structural feature of what self-reference means.

Stegemann (2026) showed that the hard problem arises only for systems satisfying the three infrastructure conditions, and that its specific difficulty, the unbridgeability of the explanatory gap, depends on the reflexive structure being present on the physical side. The present thesis completes this argument: the gap is not between the physical and the non-physical, but between two epistemic positions relative to the same physical state. Its unbridgeability is not a sign of an ontological divide but of the logical structure of self-reference, which cannot be fully circumscribed from outside.

4. Distinction from Classical Identity Theories

4.1 Type Identity and Its Failure

The classical type-identity theory (Place 1956; Smart 1959; Armstrong 1968) identified mental state types with neural state types: pain is identical with C-fibre firing, and so forth. This theory failed primarily because of the multiple realizability argument (Putnam 1967): the same mental state type can be realized by different physical state types across different creatures and systems. If pain is identical with C-fibre firing, then a creature without C-fibres cannot be in pain, which seems empirically false and conceptually arbitrary.

The present thesis avoids this failure because it identifies experience not with a physical *substrate type* but with a physical *structure type*. The relevant identity is between experience and self-reference of type *ac*, and this structural type is multiply realizable. Any system that achieves autocatalytic closure and satisfies the three infrastructure conditions instantiates Π_{ac} , regardless of its material composition. The identity is substrate-neutral: it holds for carbon-based biological systems, and in principle for any system that achieves the relevant structural organization. This makes the thesis compatible with the multiple realizability of mental states while maintaining a strict identity.

4.2 Token Identity and Its Limits

Token identity theories (Davidson 1970) identify individual mental events with individual physical events without committing to type-level identities. This avoids multiple realizability but at the cost of explanatory power: token identity tells us that this particular experience is this particular physical event, but gives no general account of why any physical event is experiential. The present thesis operates at the type level but with structural rather than substrate-based type individuation. It gives a general account: a physical state is experiential if and only if it is a state of type Π_{ac} , and this is so because the concept of experience, properly analyzed, just means a state of this type described from within.

4.3 The Structural Identity Thesis

The present thesis is a *structural identity thesis*: experience is identical with a specific structural property of physical systems, not with any specific physical substrate. Three features distinguish it from earlier identity theories. First, the structural type is defined by relational and dynamical properties, autocatalytic closure, non-factorizability, attractor-based self-reference, not by intrinsic properties of physical components. Second, the identity is not an empirical discovery but a conceptual one: it follows from the analysis of what ‘experience’ means, not from finding the physical correlate of an independently characterized phenomenon. Third, the identity entails a specific form of the two-vocabulary thesis: the phenomenal vocabulary and the physical vocabulary are not intertranslatable, but they are not descriptions of different facts, they are descriptions of the same fact from epistemically non-equivalent positions that are structurally related by the self-reference of the system in question.

5. Objections

5.1 The Knowledge Argument

Jackson’s (1982) knowledge argument presents Mary, a scientist who knows all physical facts about colour vision but has never seen red. When Mary leaves her black-and-white room and sees red for the first time, she seems to learn something new: what red looks like. If she learns something new, then what she learns was not contained in the physical facts she already knew, and therefore there are non-physical facts.

The identity thesis responds at two levels. At the first level: what Mary learns is not a new fact but a new mode of access to an old fact. She already knew the physical state Π_{ac} in its third-

person description. What she acquires on seeing red is the first-person access to that same state, the ability to instantiate it herself and thereby occupy the epistemic position from which the phenomenal description is given. This is an epistemic acquisition, not an ontological discovery. Knowing all the physical facts about swimming does not give you the ability to swim; acquiring the ability is not learning a new fact about swimming. Mary's case is analogous: she acquires an ability, not a new propositional item of knowledge.

At the second level: the knowledge argument assumes that a complete physical description is given in the third-person vocabulary. But the identity thesis implies that the first-person description is also a physical description, it is the description that the system gives of its own state when it instantiates Π_{ac} . Mary's pre-release knowledge was incomplete not because it lacked non-physical facts, but because it lacked the first-person physical description. Once she sees red, she has both descriptions of the same state. No non-physical residue remains.

5.2 The Zombie Argument

Chalmers' (1996) conceivability argument holds that philosophical zombies, physical duplicates of conscious beings that lack experience, are conceivable, and that conceivability is evidence for metaphysical possibility. If zombies are possible, then experience is not identical with any physical state.

The identity thesis responds: zombies are not conceivable once the identity is understood. A physical duplicate of a system that instantiates Π_{ac} is, by definition, a system that also instantiates Π_{ac} , and therefore a system that has experience, since experience just is that state. The apparent conceivability of zombies rests on imagining a physical duplicate while withholding the phenomenal description. But this is only coherent if the phenomenal description refers to something beyond the physical state. Once the identity is accepted, the withholding is incoherent: it is like imagining a molecule-for-molecule duplicate of water that is not wet. The intuitive pull of the zombie thought experiment is real, but it is a symptom of the two-vocabulary illusion, not evidence against the identity.

5.3 The Triviality Objection

A different objection: if the identity thesis is a conceptual rather than empirical claim, it appears to be trivially true, a matter of definition rather than discovery. And if it is merely definitional, it does not explain anything.

The response is that the distinction between conceptual and trivial does not hold here. The thesis is conceptual in the sense that it follows from an analysis of what ‘experience’ means, but this analysis is substantive and revisionary. It overturns the received view that experience is defined by its felt quality independently of its physical realization. The revisionary analysis has content: it claims that the received view embeds a false presupposition, and that once the presupposition is removed, the identity becomes visible. This is not trivially true; it is contested. Moreover, the thesis is not explanatorily empty: it implies a precise criterion for which physical systems have experience, generates empirical predictions about the boundaries of experience, and closes the explanatory gap by reidentifying rather than bridging it. That is not a small explanatory achievement.

6. Consequences for the Scientific Study of Consciousness

6.1 The Research Programme

If the identity thesis is correct, the scientific study of consciousness is not the study of how physical states produce experience, it is the study of which physical states are of type Π_{ac} , and how this structural type is instantiated across different biological and potentially artificial systems. This reorients the research programme in a specific way. The question is no longer: given that this neural state is occurring, what experience does it generate? The question is: does this system satisfy the structural conditions for Π_{ac} , and if so, which of its states are of that type?

The three infrastructure conditions identified in Stegemann (2026) provide the operational criteria. A system satisfying autocatalytic closure, criticality, and reflexive attractor structure is, in virtue of those properties, a system that has experience. The measurement of these properties, through perturbation response kinetics, metabolic signatures, and response to chronic disturbances, is the measurement of experience, not merely its correlates. There are no correlates: there is only the structural state, described from two epistemic positions.

6.2 Artificial Consciousness

The structural identity thesis has direct implications for the question of artificial consciousness. A system has experience if and only if it instantiates Π_{ac} . Whether it is biological, artificial, or of some other physical composition is irrelevant to the identity. What matters is whether it achieves autocatalytic closure, the endogenous generation and maintenance of its own self-

reference structure. Current artificial systems, including large language models, do not satisfy this condition: their self-referential properties, such as they are, are not endogenously generated but externally imposed through training and architecture. They are the computational equivalent of a superconductor in a cryostat: their relevant structure depends on external maintenance, not internal autocatalytic dynamics.

This does not foreclose the possibility of artificial experience in principle. A system that achieves genuine autocatalytic closure, that actively generates and maintains the conditions of its own self-reference, would, on the present thesis, have experience. The question for artificial consciousness research is therefore not whether silicon can be conscious, but whether any engineered system can achieve autocatalytic closure in the relevant sense.

6.3 The Boundary of Experience

The identity thesis provides a principled boundary for experience that neither panpsychism nor eliminativism can provide. Panpsychism attributes experience to all physical systems, which dissolves the concept. Eliminativism denies experience altogether, which is self-refuting, since the denial is itself an experience. The structural identity thesis draws the boundary at Π_{ac} : experience is present exactly where, and to the extent that, a system instantiates endogenously generated, self-sustaining self-reference. Below this threshold, the concept of experience is inapplicable, not because experience is absent from a thing that should have it, but because the formal condition for the application of the concept is not met.

7. Conclusion

The present article has argued for a physical identity principle for experience: $\begin{matrix} C \\ \mathcal{E} \end{matrix} \equiv \Pi_{ac}$. Experience is identical with endogenously generated, self-sustaining self-reference of the structural type that autocatalytic systems necessarily instantiate when the infrastructure conditions are satisfied. This identity is not an empirical discovery but a conceptual one: it follows from the analysis of what the term ‘experience’ means when freed from the generation assumption that has driven the hard problem.

The hard problem of consciousness is not solved by this principle. It is eliminated. The question ‘why do physical processes give rise to experience?’ rests on the false presupposition that experience is something over and above the physical state. Once the presupposition is removed, the question dissolves. What remains is the substantive scientific and philosophical task of

specifying which physical systems satisfy the structural conditions for Π_{ac} , how those conditions are implemented, and what follows from the identity for the study of mind, artificial systems, and the boundaries of experience.

The compact form of the principle is:

$$S \in Sys_{ac} \Rightarrow \mathcal{E}(S)$$

This has the logical structure of a physical law: given the conditions, the consequence is necessary. Not probable. Not expected. Necessary, because the consequence is not something added to the conditions but contained in them, as water is contained in H₂O, and as experience is contained in Π_{ac} .

References

- Armstrong, D. M. (1968). *A materialist theory of the mind*. Routledge.
- Chalmers, D. J. (1995). Facing up to the problem of consciousness. *Journal of Consciousness Studies*, 2(3), 200–219.
- Chalmers, D. J. (1996). *The conscious mind: In search of a fundamental theory*. Oxford University Press.
- Davidson, D. (1970). Mental events. In L. Foster & J. W. Swanson (Eds.), *Experience and theory* (pp. 79–101). University of Massachusetts Press.
- Jackson, F. (1982). Epiphenomenal qualia. *Philosophical Quarterly*, 32(127), 127–136.
- Levine, J. (1983). Materialism and qualia: The explanatory gap. *Pacific Philosophical Quarterly*, 64(4), 354–361.
- Place, U. T. (1956). Is consciousness a brain process? *British Journal of Psychology*, 47(1), 44–50.
- Putnam, H. (1967). Psychological predicates. In W. H. Capitan & D. D. Merrill (Eds.), *Art, mind, and religion* (pp. 37–48). University of Pittsburgh Press.
- Ryle, G. (1949). *The concept of mind*. Hutchinson.
- Smart, J. J. C. (1959). Sensations and brain processes. *Philosophical Review*, 68(2), 141–156.

Stegemann, W. (2026). The Physical Infrastructure of Experience. Zenodo preprint.
<https://doi.org/10.5281/zenodo.18800803>

Wittgenstein, L. (1953). *Philosophical investigations*. Blackwell.