Attractors Mistaken for Essences: Closing the Experiential Pathway to Essentialism

Anonymous

Abstract

Essentialism rests on two epistemological pillars: a priori arguments for the necessity of essences, and an experiential pathway—the vivid phenomenological sense that things possess necessary, mind-independent, intrinsic natures. Anti-essentialist critiques from Quine and Buddhist *śūnyatā* doctrine have largely dismantled the first pillar. This paper closes the second. Drawing on the predictive coding framework, we argue that hierarchical generative networks operating under prediction-error minimisation produce high-stability attractors whose functional profile—stability, cross-contextual consistency, independence from will, resistance to revision—matches exactly what the perception of a mind-independent essential nature would produce. From within the network's closed feedback loop, the two are structurally indistinguishable: no internal test can mark the difference. The phenomenological sense of essence therefore provides no independent evidence for the existence of essences. Husserl's *Wesensschau*, the strongest version of the experiential pathway, is shown to probe attractor-basin boundaries rather than disclose mind-independent essences: its methodological validity is preserved, but its status as an independent source of knowledge about essences is undercut. The experiential pathway from phenomenology to metaphysics is closed.

Keywords: essentialism; predictive coding; attractor; essentialist phenomenology; Husserl; *Wesensschau*; śūnyatā

1 Introduction

Essentialism—the view that things possess necessary, mind-independent, intrinsic natures—is not merely a metaphysical thesis. It has an epistemological pillar: we seem to *experience* essences. Water does not merely happen to be H_2O; it feels as though it *must* be. The law of non-contradiction does not merely hold; it feels as though it *could not fail* to hold. This experiential dimension is not incidental to essentialism but central to it. Kripke's (1980) case for metaphysical necessity has independent philosophical structure—rigid designation, causal reference theory—but at critical junctures it draws its force from thought experiments whose persuasiveness depends on the phenomenological compellingness of essentialist judgements. Husserl (1970; 1983) goes further: his *Wesensschau* (essential intuition) treats this phenomenological sense of necessity as a genuine and independent source of knowledge about essences, not a fallible inference but an intuitive fulfilment in which essential structures are directly given.

Anti-essentialist critiques have targeted the argumentative foundations of essentialism with considerable success. Quine's (1953) attack on the analytic-synthetic distinction undermines the notion that any truth is necessary in the robust logical sense. Buddhist *śūnyatā* doctrine denies that phenomena possess *svabhāva*—intrinsic, self-sufficient existence—on the grounds that everything arises through dependent origination (Nāgārjuna, 1995). These critiques operate at the level of *justification*: they dismantle the arguments for essentialism. But they leave the experiential pathway intact. As long as the phenomenological sense of essence can be reasonably interpreted as a form of epistemic contact with the world's essential structure—as long as the experiential pathway from phenomenology to metaphysics remains open—essentialism retains a fallback position. Quine can deny analyticity, but the phenomenology of logical necessity persists; the Buddhist philosopher can deny *svabhāva*, yet essentialist experience remains structurally unshaken. Buddhist soteriology itself registers this point: intellectual understanding of *śūnyatā* does not dissolve essentialist experience—precisely what one would expect if the phenomenology has a structural source independent of philosophical argument.[1]

[1]Scientific essentialism (Ellis, 2001) argues for essences via inference to the best explanation of the success

This paper argues that the experiential pathway is closed. Drawing on the predictive coding framework (Clark, 2013; Friston, 2010), we establish that the phenomenological sense of essence has a complete causal explanation that does not require the existence of essences: hierarchical generative networks operating under prediction-error minimisation produce high-stability attractors whose functional profile matches exactly what the perception of a mind-independent essential nature would produce. From within the network's closed feedback loop, the two are structurally indistinguishable. We term this structural tendency *attractor reification*. Because the closed loop affords no internal test that can distinguish attractor stability from mind-independent essence, the phenomenological sense of necessity—however vivid, however resistant to argument—provides no independent evidence for the existence of essences. The experiential pathway from phenomenology to metaphysics is blocked.

The closest predecessor to this proposal is Metzinger's (2003) analysis of the phenomenal self-model: the self-model functions as a *transparent representation* whose representational character is unavailable from within the system, producing the phenomenology of direct acquaintance with a self. The present account extends Metzinger's insight in two directions. First, it identifies the dynamical mechanism behind transparency: the self-model carries the phenomenological marks of mind-independence and necessity because it is a high-stability attractor, and any high-stability attractor produces that functional profile. Second, it generalises beyond the self: attractor reification is not a special property of self-modelling but the general mechanism through which generative networks produce the appearance of mind-independent essence. Natural kinds and logical necessities are products of the same mechanism, differing not in kind but in the robustness of the attractors involved.

The argument proceeds as follows. Section 2 develops the concept of attractor reification at the mechanistic level, establishing the structural indistinguishability of high-stability attractors and mind-independent essences from within the closed loop. Section 3 tests the completeness of this

of scientific practice. This is a distinct epistemological route—neither a priori argument nor phenomenological experience—and is not the target of the present paper. The argument here concerns whether the phenomenological sense of essence can serve as evidence; it does not address whether the success of science can.

explanation across two domains—natural kinds and logical necessity—that occupy opposite ends of the attractor robustness spectrum. Section 4 draws the epistemological conclusion: the experiential pathway is closed; Husserl's *Wesensschau*, the strongest version of that pathway, is shown to probe attractor boundaries rather than disclose mind-independent essences; and the persistence of essentialist phenomenology after philosophical argument is explained without appeal to essences. Section 5 considers objections. Section 6 concludes.

2 The Mechanism: Generative Networks and Attractor Reification

2.1 Predictive Coding and the Generative Model

The predictive coding framework, developed from Helmholtz's concept of "unconscious inference" (Helmholtz, 1867) and given its neural-computational formulation by Rao and Ballard (1999), proposes that the brain's fundamental mode of operation is not the passive reception of sensory information but the active generation of predictions. The brain maintains a hierarchical generative model of its environment—a probabilistic model of the causal structure that gives rise to sensory inputs. At each level of the hierarchy, higher areas send top-down predictions to lower areas; lower areas compute the discrepancy between predictions and incoming signals—prediction error—and feed this error signal upward. The system as a whole operates to minimise prediction error, updating the generative model when predictions fail (Clark, 2013, 2016).

On this account, perceptual content is the output of the generative model, not a direct encoding of sensory signals. What we perceive is the model's best hypothesis about the state of the world given current sensory evidence and prior expectations. The phenomenological character of perception—the sense that we are directly confronted with a world of objects—is the experiential correlate of a generative model successfully predicting its sensory input.

Friston's free energy principle (Friston, 2010) provides a more general formulation: any self-

organising system that maintains its integrity over time must minimise surprise—the long-run average of prediction error—and therefore must maintain a generative model of its environment. This extends the framework from neural systems to living systems more generally, suggesting that the tendency to model the world is not a contingent feature of human cognition but a structural requirement of any system that persists over time.

What the argument requires is the more modest claim that any optimisation system operating to minimise long-run prediction error will, as a dynamical consequence, develop stable attractor states in its internal representational landscape.[2]

2.2 Attractors in Generative Networks

A central feature of dynamical systems, including neural networks operating as generative models, is the existence of *attractors*: regions of state space toward which the system evolves and in which it stably resides (Prigogine and Stengers, 1984; Kauffman, 1993). An attractor is a stable configuration—once the system enters the basin of attraction, it tends to converge toward the attractor and resist perturbation away from it. In the context of predictive coding architectures, Friston (2019) has shown that the long-run behaviour of active inference systems can be characterised precisely in terms of attractor dynamics in the space of internal states and policies.

In a hierarchical predictive coding system, attractors correspond to configurations in which prediction error is minimised across the hierarchy. When the generative model successfully predicts incoming sensory signals at all levels, the system is in a stable, low-error state. This state has the characteristic properties of an attractor: it is the configuration toward which the system converges given the current input statistics; it is resistant to small perturbations (which generate error signals that drive the system back toward the attractor); and it is self-reinforcing (successful prediction increases the confidence, or *precision*, of the predictions that generate it).

[2]This dynamical-systems claim is compatible with enactivist frameworks (Varela et al., 1991; Thompson, 2007) in which attractors are constituted through organism-environment coupling history rather than imposed by an antecedently given internal model. The argument's force rests on the dynamical structure that both frameworks share: the tendency of coupled systems to develop stable attractors under repeated interaction.

Crucially, attractors in generative networks are **historically constituted**: the attractor landscape is shaped by the accumulated history of the system's interactions with its environment. What the system has reliably encountered in the past shapes what configurations are stable in the present. The generative model's priors—the standing expectations it brings to new sensory encounters—are the accumulated residue of past attractor convergences. In this sense, attractors are not fixed features of the network's architecture but dynamically constituted structures, shaped by learning and sedimented through repeated experience.

2.3 The Phenomenological Signature of Attractor Convergence

When a generative network converges on a high-stability attractor, the resulting representational state has several distinctive properties that together constitute *essentialist phenomenology*:

Stability: The representation is highly stable under perturbation. Attempts to represent the pattern differently generate immediate prediction error, which drives the system back toward the attractor configuration. This stability is experienced as the *resistance* of the thing to being thought otherwise—the sense that it could not be different.

Cross-contextual consistency: High-stability attractors are reached from many different starting points and under many different input conditions. The same attractor is converged upon whether the network approaches the pattern from one direction or another. This consistency is experienced as *objectivity*—the sense that the pattern is the same regardless of the perspective from which it is approached.

Independence from the subject's will: The network does not choose which attractors it converges on; convergence is driven by the system's dynamics in interaction with its inputs. The attractor is, in this sense, not the product of the subject's will but of the network's structure. This absence of voluntary control is experienced as *mind-independence*—the sense that the thing is there whether or not one wishes it to be.

Resistance to revision: Attractors, precisely because they are stable configurations, resist perturbation. Arguments that the attractor is "wrong" or "illusory" generate prediction error at

higher levels of the hierarchy, but this error is typically insufficient to dislodge a well-established attractor. This resistance is experienced as the *necessity* of the thing—the sense that no argument could make it otherwise.

These four properties—stability, cross-contextual consistency, mind-independence, and resistance to revision—are precisely the properties that the essentialist tradition attributes to essence. The essentialist claims that water is *necessarily* H_2O, that this is so *regardless* of what any subject thinks, that it is *the same* across all contexts, and that it is *resistant* to conceptual revision. On the generative account, these claims accurately report phenomenological properties of the representational state; but those properties are features of an attractor in a generative network. Why they are systematically mistaken for properties of a mind-independent essential nature—rather than of attractor dynamics—requires a separate mechanistic explanation: the structural indistinguishability argument of Section 2.5 and the feedback-channel analysis of Section 2.6.

2.4 Attractor Reification: Definition and Explanatory Scope

This paper argues that generative networks are structurally unable, from within the closed loop, to distinguish the functional profile of a high-stability attractor from the functional profile of perceiving a mind-independent essential nature, and therefore systematically represent the former as the latter. We term this structural tendency *attractor reification*. The mechanism establishes that essentialist phenomenology has a complete alternative causal source—one that requires no appeal to the existence of essences. Why this indistinguishability is structural is established in Section 2.5; the specific mechanism of outward projection is established in Section 2.6. This section delimits the explanatory target.

The account explains the *functional structure* of essentialist phenomenology—the operational profile of stability, cross-contextual consistency, independence from the subject's will, and resistance to revision—and does not claim to resolve the hard problem of consciousness.[3] The

[3] "Closed loop" in this paper refers to the architectural closure of the generative-feedback cycle, not McGinn's (1989) "cognitive closure" naming an in-principle epistemic limit on solving the mind-body problem.

epistemological consequences of this mechanistic account—for the evidential standing of essentialist phenomenology—are drawn in Section 4.

2.5 The Structural Source of Indistinguishability

Why is this functional indistinguishability structural? The argument proceeds from two premises.

Premise 1 (structural): Any generative cognitive system operates within a closed generative-feedback loop, encountering the structures it represents only *from within* this loop; it cannot occupy a neutral vantage-point outside it (Clark, 2013; Hohwy, 2013). High-stability attractors constitute, in Metzinger's sense, *transparent representations* (Metzinger, 2003): the attractor configuration is the system's settled prediction, one whose precision weight is set so high that it functions as a fixed background rather than a revisable hypothesis, and bottom-up error is systematically insufficient to revise it.

Premise 2 (conceptual): Call an attractor *maximally robust* if it appears in the asymptotic behaviour of every sufficiently expressive optimisation process operating on the relevant domain—that is, if no variation in architecture, initialisation, or loss function can avoid it. Now consider what "existing independently of the mind" means operationally from *within* the loop: a feature exists mind-independently just in case it persists however the system adjusts its processing, i.e., it is not eliminated by any cognitive variation. This is precisely the definition of maximal robustness. A maximally robust attractor and a mind-independent essential nature therefore exhibit the same operational profile from within the loop—persisting through every cognitive adjustment, generating strong error on any attempt at revision, presenting as background rather than hypothesis—and are *functionally indistinguishable*. This indistinguishability is not a contingent psychological fact but a *conceptual consequence* of the loop structure and the definition of maximal robustness: no functional test available from inside the loop can mark the difference.

This operational definition of mind-independence—"persistence under all cognitive variations"—adopts the strongest notion of mind-independence *assessable* from within the closed loop. Even on this most generous operationalisation, the distinction collapses. An essentialist might object

that mind-independence is a metaphysical concept—"would exist even if no minds existed"—not reducible to operational criteria. But this objection reinforces rather than undermines the argument: if mind-independence cannot be operationalised from within the loop, then *a fortiori* no phenomenological experience produced within the loop can serve as evidence for it. The operationalisation is not a redefinition; it is the correct delimitation of what evidence from within the loop can bear on. Cross-model convergence in large neural networks—where models trained on entirely different data and with different architectures converge to mutually predictive representations (Huh et al., 2024)—provides indirect empirical support for the existence of such maximally robust attractors.

The optimisation objective—prediction-error minimisation—is indifferent to *why* a representation is stable: whether its stability derives from an external regularity in the world or from the network's own dynamics makes no difference to the minimisation objective. There is no gradient in the optimisation landscape pointing toward the essence/non-essence distinction.

The argument is strictest for maximally robust attractors, where indistinguishability is absolute. But it extends to merely high-stability attractors in graded fashion: the deeper the attractor basin and the wider the range of input conditions from which it is reachable, the harder it is to find any functional signal from within the loop that marks the difference between "this is merely the network's stable configuration" and "this tracks a mind-independent fact." This gradient maps onto the phenomenological differences across domains: logical attractors approach maximal robustness—reinforced by every successful inference across every domain, their constraints deriving from domain-general mathematical structure—and generate the most compelling sense of necessity; natural-kind attractors are highly stable but depend on the input statistics of a particular physical environment, and their necessity-phenomenology is correspondingly less absolute. The robustness ordering makes a testable theoretical prediction: the phenomenological intensity of the "could not be otherwise" sense should vary in the same direction as the domain-generality of the constraints generating the attractor.

We can now give the full definition. *Attractor reification* is the structural tendency of generative

networks to represent the functional profile of high-stability attractors as the intrinsic nature of mind-independent essences—not a reasoning error, but the necessary consequence of any generative cognitive system satisfying the two premises above. The epistemological consequence—that essentialist phenomenology cannot serve as independent evidence for the existence of essences—is developed in Section 4.

2.6 Outward Projection: Misreading the Feedback Channel

The preceding section established that the difference between attractor stability and mind-independent essence is *undetectable* from within the closed loop. But this does not yet explain *outward projection*—why the network actively represents its attractor as a property of an external object, rather than registering it neutrally as a feature of its own dynamics. An explanation of outward projection cannot simply presuppose that error signals already have intentional content, since that would make the possession of intentional content the explanans rather than the explanandum.

The explanation begins from the *functional-causal history* of the system. A generative network develops through a history of *causal exchange with an environment*. Across this history, the system's internal states acquire what Dretske (Dretske, 1988) calls *natural information*: they carry information about external states because they were reliably caused by those states under learning conditions. On this broadly teleosemantic account, the functional role of prediction error is constituted by this causal history—error signals acquire their "aboutness" as the accumulated product of repeated cycles in which reducing error correlated with better environmental coupling. This is a naturalistic, non-circular account of how error signals come to have directional, world-oriented functional roles without presupposing intentionality from the outset (Millikan, 1984; Papineau, 1987).

Now consider what happens when this developmentally-grounded system encounters a high-stability attractor. The system's error signal approaches zero—the same functional state that, throughout its developmental history, reliably co-occurred with successful coupling to external regularities. The system has no internal procedure to determine whether the current near-zero error is the product of external tracking or of internal dynamical stability: both produce the same

functional outcome in the minimisation process. Because the functional role of near-zero error was constituted by the history of world-tracking, the system *defaults to the world-tracking interpretation*: the attractor is represented as a feature of the environment, not of the network's own dynamics. *Outward projection* is this default: a functional attribution formed under world-tracking conditions, applied in a case where the source of stability is internal.

This account avoids circularity. It does not say that error signals "mean" world-tracking and therefore get misinterpreted; it says that error signals *acquired a world-tracking functional role* through developmental causal history, and that this historically constituted role drives the outward projection even when the causal source of the signal has changed. The intentional content of the representation is explained, not presupposed.

This mechanism also explains why attractor reification cannot be eliminated by argument alone. Arguments cannot revise the developmentally constituted functional role of the error signal, nor can they dislodge the attractor's dynamical stability—the phenomenology persists after philosophical revision because both the attractor and the functional attribution that drives outward projection persist after the argument.

3 Completeness of the Explanation: Two Domains

The argument of Section 4 will require that the attractor-dynamical explanation of essentialist phenomenology is *complete*: that all phenomenological features attributed to the experience of essence are accounted for by attractor dynamics, leaving no residual feature that requires appeal to the actual existence of essences. This section tests that completeness across two domains—natural kinds and logical necessity—which occupy opposite ends of the attractor robustness spectrum. The argument assumes that the same attractor mechanism operates across the spectrum, with differences in robustness but not in kind; if the explanation is complete at both ends, it is complete across the spectrum.

3.1 Natural Kinds

The paradigm case of contemporary essentialism is the essential nature of natural kinds. Kripke (Kripke, 1980) and Putnam (Putnam, 1975) argued that natural kind terms like "water" and "gold" rigidly designate their referents by their essential microstructural properties: water is necessarily H_2O, not contingently so. The intuition driving this account is phenomenologically compelling: water *seems* to have a nature that it could not lack, a nature that obtains regardless of what anyone thinks about it.

On the generative account, this phenomenology is explained by the high-stability attractor the cognitive network has formed around the cluster of properties—visual, tactile, functional—that reliably co-occur in encounters with water. This attractor is converged upon from a wide range of starting conditions, generates strong prediction error when any of its core features are challenged, and is resistant to perturbation. The essentialist intuition—that water could not fail to be H_2O—is the phenomenological report of this resistance.

The generative account does not deny that water is H_2O; it explains why this truth *feels* necessary. The discovery that water is H_2O provides a scientific description of the external regularity that the attractor has reliably tracked—without establishing that this regularity constitutes an intrinsic, mind-independent *essence* in the philosophically loaded sense. The phenomenology would be exactly the same whether or not water's H_2O constitution constitutes an intrinsic essence, because the functional profile is fully explained by the attractor's stability.

Natural-kind attractors occupy a specific position in the robustness ordering. Their constraints derive from the regularities of a particular physical environment, which are universal within the physical universe but contingent relative to the space of possible environments. They are therefore highly stable but fall short of maximal robustness. All four phenomenological marks—stability, cross-contextual consistency, mind-independence, resistance to revision—are fully covered by the attractor account without residue.

3.2 Logical Necessity

The necessity of logical and mathematical truths requires a preliminary distinction. Two questions must be carefully separated: the **ontological question** (why are logical truths true, and necessarily so?) and the **phenomenological question** (why does our grasp of them feel so compellingly necessary?). The argument targets the phenomenological question *only*. It does not attempt to explain why logical truths are true; it accepts Frege's anti-psychologism in full. What it explains is the specific phenomenological character of grasping a logical truth—why that grasp feels like the direct apprehension of necessity.

On the generative account, logical and mathematical attractors are among the highest-stability attractors in the network's state space, because they are reinforced by every successful inference across every domain. They approach maximal robustness: appearing in the asymptotic behaviour of any sufficiently expressive cognitive process operating on discrete structure. This is why the phenomenology of logical necessity approaches the limiting case—they are the closest human cognizers come to a genuinely inescapable attractor.

One concession is in order: part of the reason logical attractors are so stable is presumably that logical truths are true, and cognitive systems tracking them reliably succeed. The generative account does not deny this; the tracking objection in Section 5 addresses the general case. What the account denies is that the phenomenological sense of necessity is, by itself, evidence for any particular view about the ontological question. The same functional profile would arise for any attractor of comparable stability, regardless of its ontological status.

Again, all four phenomenological marks are fully covered. Logical necessity occupies the high end of the robustness spectrum: maximally stable, maximally cross-contextual, maximally independent of will, maximally resistant to revision. The attractor account predicts this ordering and explains it without residue.

Across the two domains, the attractor-dynamical explanation covers the full spectrum of attractor robustness and accounts for all four phenomenological marks at each point. No residual phenomenological feature has been identified that would require appeal to the existence of essences.

The explanation is complete. This completeness is a necessary premise for the epistemological argument that follows.

4 The Epistemological Conclusion

4.1 Closing the Experiential Pathway

The preceding sections have established two results. First, the structural indistinguishability argument (Section 2.5): from within the closed generative-feedback loop, the functional profile of a high-stability attractor and the functional profile of perceiving a mind-independent essential nature are indistinguishable—no internal test can mark the difference. Second, the completeness of explanation (Section 3): attractor dynamics fully account for all four phenomenological marks of essentialist experience—stability, cross-contextual consistency, mind-independence, resistance to revision—across the full robustness spectrum, without residue.

Together, these results close the experiential pathway from phenomenology to metaphysics. The argument has the following structure:

1. The phenomenological sense of essence (the "it could not be otherwise" feeling, the sense of mind-independence, and so on) has a complete causal explanation in attractor dynamics that does not require the existence of mind-independent essences. (Sections 2–3)

2. From within the closed loop, no functional test can distinguish whether this phenomenological sense is produced by an attractor or by genuine contact with a mind-independent essence. (Section 2.5)

3. Therefore, the phenomenological sense of essence provides no independent evidence for the existence of essences. The experiential pathway from phenomenology to metaphysics is closed.

This argument does not presuppose that essences do not exist. It is compatible with the metaphysical possibility that essences exist and that some attractors track them. What it establishes

is that *even if* essences exist and *even if* some attractors track them, the phenomenological sense of essence cannot serve as evidence for this, because the same phenomenology would arise regardless. The argument is epistemological, not metaphysical: it concerns the evidential standing of essentialist phenomenology, not the existence or non-existence of essences.

This distinguishes the argument from a standard debunking argument. Debunking arguments in metaethics (Street, 2006) typically claim that a belief's causal origin is unrelated to the truth of what is believed, and therefore the belief is unjustified. The present argument makes a different and stronger claim: the causal source provides a *complete* alternative explanation for the phenomenology, *and* the system has no internal means of distinguishing the two sources. It is not that the causal origin is merely unrelated to truth; it is that the system is structurally unable to determine whether it is in contact with essence or with its own dynamics. The evidential pathway is not merely undercut; it is closed.

4.2 Consequences for Husserl's Essential Intuition

Husserl's *Wesensschau* (essential intuition) represents the strongest version of the experiential pathway—and therefore the most demanding test case for the argument. In the *Logical Investigations* (Husserl, 1970) and *Ideas I* (Husserl, 1983), Husserl argues that essences are not merely felt but *directly given* through a distinctive form of intuition. The method of *free variation in imagination*—systematically varying the features of a thing and attending to what remains invariant—discloses universal essences with a certainty proper to categorial cognition. For Husserl, this is genuine intuitive fulfilment, not fallible inference: the essence is "bodily given" (*leibhaft gegeben*), just as a perceived object is given in sensory perception. *Wesensschau* thus claims to be an independent source of knowledge about essences—one that does not reduce to induction, conceptual analysis, or any other mode of knowing. It claims, in short, to provide a bridge from experience to essence.

If the experiential pathway is closed, this bridge cannot bear the weight placed on it.

The argument proceeds in two steps. First, a redescription: free variation is, from the perspective of the generative network, a procedure that probes the stability boundaries of a high-stability attractor.

The subject explores how far the basin of attraction extends—varying features in imagination until the network produces strong prediction error. For a generative network, imagining a variation and perceiving a variation are both operations that generate top-down predictions; the prediction-error cost of each reveals the depth of the attractor basin. What Husserl calls the "essentially necessary"—what cannot be varied away—corresponds to the features whose removal would most severely destabilise the attractor. The "invariant" disclosed by free variation is not a Platonic essence floating outside any cognitive system but a map of the attractor's deepest basin.

Second, the epistemological consequence: the phenomenological quality that accompanies successful *Wesensschau*—the compelling sense that one is in direct contact with a necessary, mind-independent structure—is the phenomenological signature of convergence to a high-stability attractor. By the structural indistinguishability argument (Section 2.5), this signature is the same whether the attractor tracks a genuine essence or not. Therefore, the phenomenological quality of *Wesensschau* cannot serve as evidence that what is disclosed is a mind-independent essence rather than an attractor-basin boundary.

This does not render *Wesensschau* methodologically worthless. Systematic variation *does* disclose something stable about the network's cognitive organisation. But the epistemological status of what it discloses is downgraded: the invariant is invariant relative to *this type of cognitive network operating on this type of input statistics*. Whether this relative invariance also constitutes a mind-independent essential structure is a further question that *Wesensschau* itself cannot answer—because the phenomenological quality that seems to guarantee mind-independence is fully explained by attractor dynamics.

The argument respects the distinction between the transcendental and the naturalized level. Husserl's transcendental phenomenology asks what meaning-constituting structures make objects intelligible as such; the attractor account asks why encounters with high-stability configurations carry the specific phenomenological stamp of necessity. These are different questions, and the attractor account does not claim to answer the transcendental one. What it does claim is that the phenomenological quality disclosed by *Wesensschau*—the very quality that seems to bridge from

experience to essence—is not a reliable bridge. The bridge is built from attractor dynamics, and it leads back into the network rather than out toward mind-independent essences.

4.3 Why Argument Cannot Dissolve the Phenomenology

The experiential pathway is closed, but the phenomenology will not disappear. This is not a defect of the argument but a prediction of it. The argument changes the *epistemological status* of essentialist phenomenology; it does not change the *attractor dynamics* that produce it. High-stability attractors persist after the argument is accepted, and the phenomenological signature of convergence persists with them. The world does not stop appearing as though things have essential natures.

This explains an otherwise puzzling fact about the history of anti-essentialism. Quine's arguments are widely accepted, yet logical truths still feel necessary. Buddhist practitioners who intellectually grasp *śūnyatā* report that essentialist phenomenology persists until sustained meditative practice—not mere intellectual assent—changes the network's attractor landscape. In predictive-coding terms, what sustained practice changes is the *precision weighting* assigned to the meta-cognitive prediction "this stability indicates a mind-independent essence": the practitioner comes to recognise the phenomenological mark of attractor convergence for what it is, rather than automatically endorsing the essentialist reading (Siderits, 2003). Intellectual conviction can revise beliefs, but precision weights are revised through the accumulated history of attention—a distinction that explains why Buddhist soteriology insists on practice rather than argument.

The present paper closes the experiential pathway at the level of *epistemology*. Quine's critique closes it at the level of *justification*. The two are complementary: Quine shows that no argument can establish that a truth is necessary; the present paper shows that no experience can provide evidence that it is. Together, they leave essentialism without epistemological support—though the phenomenology, being a structural product of generative networks, will endure.

5 Objections and Replies

5.1 The Tracking Objection

Objection: Perhaps high-stability attractors are high-stability precisely because they track real regularities in the world. If so, the stability of the attractor is evidence of, not a substitute for, a genuine mind-independent pattern. Attractor reification, on this view, is not a mistake but the normal case of successful world-modelling.

Reply: The objection is partly correct. High-stability attractors often do track genuine regularities: the attractor for "water" is stable in part because H_2O genuinely has stable properties across contexts. The argument does not deny this. What it denies is that the phenomenological signature of attractor convergence—the sense of necessity, mind-independence, and intrinsic nature—constitutes evidence for a further, distinctively *essentialist* fact about the world. There is a gap between "this attractor tracks a genuine regularity" and "this regularity has a necessary, intrinsic, mind-independent nature." Essentialist phenomenology conflates these two claims; the argument separates them. As the structural indistinguishability argument establishes, the functional profile produced from within the generative loop is identical whether or not an attractor tracks an external essence, because the loop has no internal test for that distinction. The tracking objection shows that attractor reification need not be a global error—some attractors may well track genuine structures. But well-calibrated tracking does not vindicate the essentialist interpretation of the phenomenology, because the phenomenology provides no signal that distinguishes tracking from mere stability.

5.2 The Circularity Objection

Objection: The argument that the experiential pathway is "closed" presupposes that there are no mind-independent essences. But this is precisely what is in dispute. The argument begs the question against the essentialist.

Reply: The argument does not presuppose that there are no mind-independent essences; it is compatible with a range of positions on that metaphysical question. What it claims is that the

phenomenological sense of necessity is not, by itself, evidence for mind-independent essences, because that sense is fully explained by the dynamics of the closed generative-feedback loop, independently of whether the attractor tracks an essence or not. This is a claim about the *epistemic standing* of essentialist phenomenology, not about the metaphysics of essence. Even an essentialist should accept that phenomenology alone—without independent argument—does not settle the metaphysical question, because the phenomenology would be exactly the same whether or not essences exist.

The argument's contribution is mechanistic, not definitional: it gives a *dynamical-mechanistic* explanation of why cognitive systems produce states with this functional profile—because prediction-error minimisation necessarily carves out stable configurations in state space. Prior to this account, we could describe essentialist phenomenology (stability, resistance to revision, mind-independence) but had no account of *why* these features cluster together, *why* they arise with the specific qualitative character of necessity rather than mere familiarity, and *why* they persist under philosophical revision. The attractor account answers all three by locating their common source in the dynamical structure of generative networks.

5.3 The Perceptual Analogy

Objection: Visual perception also has a complete neural-causal explanation. From within the visual system, veridical perception and hallucination are functionally indistinguishable. Yet we do not conclude that visual experience provides no evidence for the existence of external objects. If the argument generalises to perception, it yields radical scepticism—a reductio.

Reply: The argument does not generalise to perception, because perception has something that essentialist phenomenology lacks: an independent calibration channel. We trust visual perception not because of its phenomenological quality—its vividness, its sense of direct presence—but because perception-guided action systematically succeeds: reaching for a seen object makes contact; walking toward a seen wall produces a collision. Behavioural consequences provide feedback that is independent of the phenomenological quality of the percept. This feedback loop enables calibration:

we can distinguish veridical perception from hallucination not by inspecting phenomenological quality alone, but by checking whether action guided by the percept produces the expected outcome.

No analogous calibration channel exists for essentialist phenomenology. There is no action one can perform whose success or failure would distinguish "this attractor tracks a genuine mind-independent essence" from "this attractor is merely very stable." The phenomenological quality—the sense of necessity, the sense of mind-independence—is all there is. When that quality is shown to have a complete alternative causal source and the system has no independent means of calibration, the evidential pathway is closed. The perceptual case does not meet this second condition; the essentialist case does.

6 Conclusion

This paper has argued that the experiential pathway from essentialist phenomenology to essentialist metaphysics is closed. The mechanism is *attractor reification*: generative networks operating under prediction-error minimisation produce high-stability attractors whose functional profile—stability, cross-contextual consistency, independence from will, resistance to revision—matches exactly what the perception of mind-independent essential natures would produce. From within the closed feedback loop, the two are structurally indistinguishable. Because the loop affords no internal test to mark the difference, the phenomenological sense of essence provides no independent evidence for the existence of essences.

This conclusion has three consequences. First, the *ubiquity* of essentialist phenomenology is explained: any system operating under prediction-error minimisation will produce high-stability attractors and therefore essentialist phenomenology, whether or not essences exist. Second, the *persistence* of that phenomenology in the face of anti-essentialist argument is explained: the phenomenology is produced by attractor dynamics, not by arguments, so refutation cannot dislodge it. Third, Husserl's *Wesensschau*—the most developed version of the experiential pathway—is shown to probe attractor-basin boundaries rather than disclose mind-independent essences: methodologically

valid as a way of mapping cognitive organisation, but unable to serve as an independent source of knowledge about essences.

A separate argument establishes an asymmetry in ontological commitment: determinate structure requires justification; maximal indeterminacy does not (Author, manuscript).[4] That argument closes what might be called the *a priori* pathway to essentialism—the assumption that determinacy is the default condition of being. The present paper closes the *experiential* pathway. Together, they leave essentialism without epistemological support: neither a priori argument nor experiential evidence can establish that things possess necessary, mind-independent, intrinsic natures.

The most parsimonious ontological picture that emerges is one in which essentialist phenomenology is a structural product of generative networks—not a window onto mind-independent essential natures, but the characteristic cognitive experience of high-stability attractors converging within a closed feedback loop. Whether any attractors track genuine structures in the world remains an open question of metaphysics and epistemology. What is no longer open is whether the phenomenological sense of essence can answer it. It cannot.

The deepest implication concerns the relationship between cognitive architecture and philosophical understanding. The persistence of essentialist phenomenology is not a failure of reasoning that better philosophy could overcome; it is a structural feature of the kind of minds we are. Understanding this does not dissolve the phenomenology—the world does not stop appearing as though things have essential natures—but it changes the epistemic weight we should assign to that appearance. Learning to hold this gap open—to undergo the appearance of essence while understanding its generative origin—may be the closest that cognitive systems of our kind can come to what the Buddhist tradition describes as insight into *śūnyatā*.

[4] Citation removed for double-blind review.

Statements and Declarations

Competing Interests: The author declares no competing financial or non-financial interests related to this work.

Funding: No funding was received for conducting this study.

Use of AI assistance: The drafting and revision of this manuscript was assisted by Claude (Anthropic), a large language model, under the author's direction. All intellectual content, arguments, and interpretive decisions are the author's own.

References

Andy Clark. Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behav Brain Sci*, 36(3):181–204, 2013. doi: 10.1017/S0140525X12000477.

Andy Clark. *Surfing Uncertainty: Prediction, Action, and the Embodied Mind*. Oxford University Press, New York, 2016.

Fred Dretske. *Explaining Behavior: Reasons in a World of Causes*. MIT Press, Cambridge, MA, 1988.

Brian D. Ellis. *Scientific Essentialism*. Cambridge University Press, Cambridge, 2001.

Karl Friston. The free-energy principle: a unified brain theory? *Nat Rev Neurosci*, 11(2):127–138, 2010. doi: 10.1038/nrn2787.

Karl Friston. A free energy principle for a particular physics, 2019.

Hermann von Helmholtz. *Handbuch der Physiologischen Optik*, volume 3. Voss, Leipzig, 1867. Vol. 3 contains the theory of unconscious inference; the complete work was published in three volumes, 1856–1867.

Jakob Hohwy. *The Predictive Mind*. Oxford University Press, Oxford, 2013.

Minyoung Huh, Brian Cheung, Tongzhou Wang, and Phillip Isola. Position: The Platonic Representation Hypothesis. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *PMLR*, pages 20617–20642, 2024.

Edmund Husserl. *Logical Investigations*. Routledge, London, 1970. Translated by J. N. Findlay; originally published 1900–1901.

Edmund Husserl. *Ideas Pertaining to a Pure Phenomenology and to a Phenomenological Philosophy: First Book*. Martinus Nijhoff, The Hague, 1983. Translated by F. Kersten; originally published as *Ideen zu einer reinen Phänomenologie* 1913.

Stuart A. Kauffman. *The Origins of Order: Self-Organization and Selection in Evolution*. Oxford University Press, New York, 1993.

Saul A. Kripke. *Naming and Necessity*. Harvard University Press, Cambridge, MA, 1980.

Colin McGinn. Can we solve the mind-body problem? *Mind*, 98(391):349–366, 1989. doi: 10.1093/mind/XCVIII.391.349.

Thomas Metzinger. *Being No One: The Self-Model Theory of Subjectivity*. MIT Press, Cambridge, MA, 2003.

Ruth Garrett Millikan. *Language, Thought, and Other Biological Categories*. MIT Press, Cambridge, MA, 1984.

Nāgārjuna. *The Fundamental Wisdom of the Middle Way: Mūlamadhyamakakārikā*. Oxford University Press, Oxford, 1995. Translated with commentary by Jay L. Garfield.

David Papineau. *Reality and Representation*. Blackwell, Oxford, 1987.

Ilya Prigogine and Isabelle Stengers. *Order Out of Chaos: Man's New Dialogue with Nature*. Bantam Books, New York, 1984.

Hilary Putnam. The meaning of "meaning". *Minnesota Stud Philos Sci*, 7:131–193, 1975.

Willard Van Orman Quine. *From a Logical Point of View*. Harvard University Press, Cambridge, MA, 1953. Contains "Two Dogmas of Empiricism".

Rajesh P. N. Rao and Dana H. Ballard. Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nat Neurosci*, 2(1):79–87, 1999. doi: 10.1038/4580.

Mark Siderits. *Personal Identity and Buddhist Philosophy: Empty Persons*. Ashgate, Aldershot, 2003.

Sharon Street. A Darwinian dilemma for realist theories of value. *Philosophical Studies*, 127(1): 109–166, 2006.

Evan Thompson. *Mind in Life: Biology, Phenomenology, and the Sciences of Mind*. Harvard University Press, Cambridge, MA, 2007.

Francisco J. Varela, Evan Thompson, and Eleanor Rosch. *The Embodied Mind: Cognitive Science and Human Experience*. MIT Press, Cambridge, MA, 1991.