The Real Illusion of Consciousness

On Retiring a Category That Never Earned Its Place

Working Paper — 2026

shakyfoundation.com — Topic 002

Before we begin

Pain hurts. Grief is heavy. The face of someone you love produces something specific and irreplaceable when you see it. The moment before sleep, when the day's noise fades and something quieter takes over — that is real. The reader who opened this paper with a flinch of alarm at its title — that flinch was real. The act of noticing it was real. The decision to keep reading was real.

Nothing in this paper denies any of that.

What follows is not an argument that inner life is an illusion. It is not a claim that you are a machine and your experiences are mere computation. It is not a denial that suffering matters, that love is significant, or that the self is real. If at any point the argument seems to be heading in one of those directions, something has gone wrong — return to this paragraph. The phenomena are safe.

The argument is narrower than it sounds and considerably less frightening than its title suggests. It is this: the word "consciousness," treated as the name of a single unified thing, does not pick out anything that science has been able to find, define, or measure — after three centuries of trying. The word bundles together a collection of real phenomena — attention, self-awareness, emotion, the sense of presence, the felt quality of experience — and implies that they constitute a single natural kind requiring a single explanation. They do not. They are many things, each of them real, and each of them better understood when studied on its own terms rather than forced under a label that has never been made precise.

Removing the label does not remove what was being pointed at. It removes the obstacle that has prevented clear thinking about what was being pointed at. That is the entire argument.

Abstract

This paper argues that "consciousness," treated as a unified ontological category, is not a discovery but a posit — and that once the real phenomena are disaggregated, the posit does no explanatory work that more precise vocabulary cannot do better. What dissolves is not the phenomena but the unnecessary extra category imposed on top of them.

The paper advances two related but distinct claims. The first is a taxonomy claim: "consciousness" does not pick out a unified natural kind; the phenomena grouped under the term are too heterogeneous to share a common explanatory structure, and they dissociate readily under brain damage, anaesthesia, and altered states. The second is an anti-residue claim: once those phenomena are accounted for by the processing-system account, nothing clearly remains that would justify positing consciousness as an additional ontological category. A critic might grant the first while contesting the second, so the paper argues for them separately.

This position is importantly distinct from illusionism (Frankish) and from the position of Dennett, who correctly strips qualia of their metaphysical weight and rejects the Cartesian theater, but retains the word "consciousness" and with it the assumption of a unified target phenomenon. This paper denies less than illusionism and goes further than Dennett. The claim is not that inner states are illusory, but that the category "consciousness" does not pick out a coherent natural kind on top of the real processes it bundles together. The phenomena are real; the unifying label is unwarranted. This account is more conservative with respect to lived experience than illusionism, not less.

A processing-system account is constructed from defensible primitives: input, internal state, state-change, self-modeling, and output. All phenomena standardly attributed to consciousness are recovered within this account, located more precisely than before. The term "consciousness" is shown to be eliminable without loss — and its elimination opens a more tractable research programme than its retention. In the terminology of the eliminativism literature, this is discourse eliminativism rather than entity eliminativism: the things people point to when they use the word are real; the grouping of them under a single term is not.

The account has consequences beyond philosophy of mind. Once the binary conscious/not-conscious collapses into a spectrum of aversive-state capacity and self-modeling depth, the questions of how we treat animals and AI systems become more tractable — and more urgent — than the consciousness framework ever allowed.

Keywords: consciousness, category elimination, discourse eliminativism, processing-system account, philosophy of mind, qualia, illusionism, Cartesian dualism, suffering, self-reference, internal states

---

Part I: Preamble

Why this matters

Three centuries after Descartes, philosophy of mind remains deeply shaped by a problem his framework sharpened and helped stabilise. The "hard problem of consciousness" — why physical processes give rise to subjective experience — is treated as one of the deepest unsolved questions in science and philosophy. Entire research programmes, journals, and careers are built around it.

This paper argues the problem is not hard. It is malformed. It was generated by an unexamined assumption introduced in the 17th century and never seriously questioned since. Remove the assumption and the problem does not get solved — it disappears. What remains are real, tractable questions about how processing systems work, how

they model themselves, and how they enter and sustain internal states. Those questions have answers. They are being answered. They did not need the word "consciousness" to get started, and they will not miss it when it is gone.

This matters beyond philosophy. How we define consciousness determines how we treat animals, how we think about AI systems, and how we draw the boundaries of moral consideration. A category that cannot be defined, cannot be measured, and cannot be agreed upon is doing ethical work it is not qualified to do. Replacing it with tractable concepts does not shrink our moral concern — it grounds it.

What this paper argues — and what it does not

The central claim is that "consciousness" does not name a natural kind. It is not a scientific term that has proved difficult to define. It is a folk category that has proved impossible to define because it was never tracking a single thing to begin with. It bundles together attention, self-modeling, affect, global access, agency, and the sense of presence — real phenomena, all of them — under a label that implies they constitute a unified kind. They do not.

Removing the label leaves all of those phenomena intact. Nothing in lived experience is denied. Pain still hurts. Emotions are still real. The self still coheres. The denial is of the category imposed on top of those things, not of the things themselves.

The title is meant precisely. Dennett called consciousness an illusion. This paper argues he was right that there is an illusion — but wrong about what it is. The illusion is not experience. It is the false unity of the category.

What this paper is not

This position is easily misread, so it is worth being exact about what is and is not being claimed.

This is not illusionism. Frankish argues that phenomenal consciousness seems to exist but doesn't — that we are systematically deceived about our own inner qualitative states, that there are in his words "no inner sounds, smells, tastes and pains" as qualitative items. This paper makes no such claim. Inner states are real. Pain is real. The claim is narrower: the category "consciousness" does not pick out a coherent natural kind on top of those real states. We are not being asked to accept that experience is an illusion. We are being asked to stop treating a loose collection of real things as though they add up to a unified metaphysical kind.

This is not Dennett's position. Dennett correctly strips the magic from consciousness — rejects the Cartesian theater, rejects qualia as traditionally conceived, insists that explaining the functions is explaining the phenomenon. But Dennett keeps the word and with it the assumption that there is a unified target phenomenon being explained. That assumption is what this paper rejects. Dennett demystifies consciousness. This paper retires it.

This is not skepticism about experience. Anyone who suspects this account explains away inner life has it backwards. Illusionism explains away inner life. This account affirms it — locates it precisely, gives it causal reality, and stops pretending it needs a mysterious extra category to be taken seriously.

The landscape, stated as a table:

Position	The phenomena	The category
Realism (Chalmers)	Real	Real and irreducible
Illusionism (Frankish)	Illusory	Real as a target of the illusion
Dennett	Real	Real but deflated
This paper	**Real**	**Unwarranted — bundles real things under a false unity**

The Moorean objection — and why it helps rather than hinders

The most common gut-level resistance to any critical account of consciousness takes roughly this form: "Whatever arguments you offer, I am more certain that I am conscious than I am that your premises are correct." Chalmers calls this the Moorean argument against illusionism. It is a real objection — but it lands against illusionism, not against this paper.

The Moorean argument works by asserting that the reality of experience is more certain than any theory that denies it. This paper does not deny the reality of experience. It denies that "consciousness" is the right name for a unified natural kind that encompasses all of it. The certainty of inner experience is fully compatible with the claim that the category bundling those experiences together is scientifically unwarranted. Anyone who feels the force of the Moorean intuition — "I know I am experiencing something, and no argument can shake that" — should notice that this paper agrees completely. The Moorean intuition is an ally of this account, not an objection to it.

A note on intellectual neighbours

One near-neighbour deserves acknowledgment before the argument begins. Jacy Anthis has proposed what he calls "consciousness semanticism," which distinguishes "consciousness-as-self-reference" from "consciousness-as-property" and argues the latter should be eliminated while the former survives. This is structurally similar to the position taken here, and the overlap is genuine. The difference is one of framing and foundation: Anthis approaches the question primarily through the lens of semantic analysis; this paper approaches it through the lens of scientific ontology and the burden of proof. The conclusions converge more than they diverge.

Structure of the argument

The paper proceeds in order of logical dependence.

Part II establishes what is not lost — before the argument begins, the phenomena should be known to be safe. Part III traces how the assumption entered — Descartes, the mind-body split, and why every subsequent attempt to escape it stopped short. Part IV makes the burden of proof explicit — three centuries of failed definition is not bad luck, it is diagnostic. Part V constructs the positive account from defensible primitives. Part VI does the recovery accounting — each phenomenon walked through in detail. Part VII draws out the implications for animals, AI, and ethics. Part VIII addresses the objections

directly. Part IX states the open questions honestly. The coda says what has been built and what remains.

Part II: Nothing Is Lost

A paper with the word "illusion" in its title and "consciousness" as its target invites a particular kind of alarm. It is reasonable to wonder whether what follows will deny the reality of their pain, dismiss their grief as mere computation, or reduce their sense of being someone to a convenient fiction. That reaction is worth pausing on — not to dismiss it, but because the act of stepping back to observe it is itself instructive. Something just noticed its own response, evaluated it, and decided to keep reading anyway. That capacity — to model oneself, to observe one's own reactions, to maintain a perspective across time — is real, and nothing in this paper touches it.

The argument that follows is not an attack on inner life. It is an argument that inner life has been mislabelled, and that the mislabelling has caused more confusion than clarity. Removing the label does not remove what was being pointed at.

This Part exists to make that clear before the philosophical work begins. No one should have to wait until Part V to learn whether their inner life survives the argument. It does. Here is how.

2.1 The phenomena are preserved

Consider the things that matter most when people insist that consciousness is real.

Pain. The sharp, unmistakable signal that something is wrong. The way it overrides other concerns, redirects attention, and demands a response. On the processing-system account, pain is an aversive internal state with genuine causal power — it reshapes the organism's processing toward escape, avoidance, and relief. That causal power is not a metaphor. It is not a "mere correlate." It is what pain *is*: a state that drives the system, powerfully and urgently, away from what is harming it. Nothing about that is diminished by describing it precisely.

Grief. The sustained, heavy reconfiguration that follows a loss the self-model has not yet integrated. Grief is not a single feeling — it is a cascade of internal states: the searching for what is absent, the repeated collision with the fact of loss, the slow restructuring of predictions and plans that had included the person who is gone. Each of these is real, causally efficacious, and located precisely in the processing-system account. Calling it "an aspect of consciousness" adds nothing to this description. It subtracts precision.

Love. The recognition of a specific face, a specific voice, a specific presence — and the way that recognition produces something powerful and particular in the organism. The heart rate shifts. Attention narrows. The internal state reconfigures toward approach, toward care, toward a specific kind of vulnerability. This is not mere computation. It is the deepest kind of internal state a social organism can sustain. The processing-system account does not explain love away. It says what love does, how it works, and why it matters — which is more than the word "consciousness" has ever managed to do.

The quality of a moment. A morning in early spring when the light is a particular way. The difference between hearing a piece of music for the first time and hearing it again

years later when it is saturated with memory. The specific character of waking from a dream that has not quite faded. These are not candidates for elimination. They are the starting point. The processing-system account does not deny that moments have character. It says that the character is constituted by the internal states — perceptual, affective, mnemonic, self-locating — that are active in the organism at that moment. The character is real. It does not need a metaphysical wrapper to be taken seriously.

Pleasure, mood, imagination, introspection, dreaming — all of these survive. They are recovered as what they always were: real activities of a real processing system, now described without the overhead of a category that was never earning its keep.

2.2 The self survives

Of all the things people fear losing when "consciousness" is questioned, the self is perhaps the most charged. The suspicion is that if consciousness goes, the self goes with it — that without some inner light of awareness, there is no one home, no subject, no "I" that persists.

This paper argues the opposite. The self not only survives the retirement of "consciousness" — it gets a more precise and defensible account than consciousness-based frameworks typically provide.

A self, on the processing-system account, is a system that maintains a model of itself: that refers to itself in its processing, distinguishes itself from its environment, tracks its own states across time, and uses that self-model to plan, predict, and act. This is not a thin or deflationary notion of selfhood. It is a rich functional description of what a self actually does — and it is demonstrable, graduated, and does not require invoking anything mysterious to get started.

Whoever noticed their own alarm at this paper's title, stepped back, evaluated it, and decided to continue — that is exactly what a self does. The self-model is running. It is working. Nothing about that process requires the word "consciousness" to be real.

What the processing-system account adds is precision about degrees. A very simple organism has a rudimentary form of self-boundary — it distinguishes self from non-self for immune purposes. A more complex organism has richer self-modeling — it can represent its own emotional states, anticipate its own responses, and adjust its behaviour based on a model of itself as an agent. A human being has the most elaborate self-model known: one that includes a narrative of personal history, a theory of its own mind, and the capacity to step back and observe its own reactions — as you may have noticed yourself doing throughout this section.

None of these capacities requires the word "consciousness" to be real, and none of them is threatened by its retirement.

2.3 What illusionism would actually ask of you

It is worth being precise about what this paper is *not* asking, because the contrast with illusionism matters.

Illusionism, as defended by Keith Frankish, holds that phenomenal consciousness — the inner qualitative character of experience, what it is like to feel pain or see red — does not actually exist. We are, on this view, systematically deceived about our own inner

states. There are, in Frankish's explicit formulation, no inner sounds, smells, tastes, or pains as qualitative items. What we take to be the felt quality of experience is an illusion generated by the brain.

That is a demanding position. It asks you to accept that when pain hurts, the hurting is not quite what it seems — that the felt quality is somehow not real even as the functional state is. Many find this not just counterintuitive but incoherent.

This paper asks for something considerably less dramatic:

1. Pain is real and causally powerful. ✓
2. Emotions are real internal states with real effects. ✓
3. The self is a real structural feature of the organism. ✓
4. The specific character of any given moment — its quality, its texture, its feel — is constituted by the internal states active in the organism at that moment. ✓
5. The word "consciousness," which bundles all of these together and implies they constitute a unified natural kind, is not scientifically warranted — and retiring the word does not touch any of the above. ← This is the claim.

That is the entire distance between your current position and this paper's conclusion. If points 1 through 4 are granted — and they should be, because they are just descriptions of what the phenomena are — then the question is only whether the word "consciousness" is doing useful work on top of them. The paper argues it is not.

2.4 A promise

The paper will make its case in the Parts that follow — tracing the history of the assumption (Part III), examining the failure of three centuries of definition (Part IV), constructing the positive account (Part V), recovering every phenomenon in detail (Part VI), drawing out the consequences (Part VII), and addressing the objections directly (Part VIII).

Throughout all of that: nothing that is real will be denied. If the argument ever seems to be denying something real, the author has failed, and you should hold the author to the commitment made here.

What is at stake is a word, a grouping, a category — not the phenomena the category was imposed on. The phenomena are safe. They were always safe. The word was the problem, not the things it was trying to name.

Part III: How the Assumption Entered

The hard problem of consciousness in its modern form is a post-Cartesian construction. Before Descartes stabilised the framework in which it became thinkable, the question "why does physical processing give rise to subjective experience?" did not exist in its current form, because the sharp divide between physical and mental that makes the question possible did not exist in the same way. Earlier thinkers — Aristotle, the Stoics, Aquinas — wrestled with perception, sensation, soul, and first-person awareness. But the specific problem of bridging an ontological gap between two categorically distinct substances was not yet the organising frame. The mind-body problem, as philosophy of

mind inherits it, is not a timeless puzzle. It is a product of a specific historical decision, made for reasons that had more to do with theology than with science, and it has been organising the field ever since.

Understanding where the assumption came from is not merely historical interest. It is prerequisite for recognising why the assumption has been so difficult to dislodge. A problem that was *created* by a framework cannot be *solved* within that framework. It can only be dissolved by stepping outside it.

3.1 Descartes and the original split

René Descartes introduced the mind/body distinction in the 1640s — not as a scientific hypothesis but as a solution to a theological problem. He needed to carve out a domain for the soul: something that could survive bodily death, stand outside the mechanical world of extended matter, and bear the kind of moral responsibility that Christian doctrine required. At the same time, he wanted to give a fully mechanical account of the body — one that explained animal behaviour, reflexes, and physiology without invoking the soul at every turn. The solution was a sharp ontological divide. Mind was *res cogitans*, thinking substance, non-extended and non-mechanical. Body was *res extensa*, extended matter operating by physical law. The two were declared categorically distinct.

The split was widely accepted — not because it solved the philosophical problem cleanly (it immediately created the interaction problem: how does a non-physical mind move a physical arm?), but because it was theologically convenient and culturally legible at a moment when both mattered enormously. It gave science the body and gave theology the soul. For a generation, everyone was satisfied — or at least, everyone who mattered to the institutions that published and promoted philosophy.

What the split also did, less visibly, was create a conceptual residue. If mind and body are categorically distinct, something is needed to name whatever it is that minds have and bodies don't. That something became "consciousness" — defined not by what it is, but by what it is not. It is not physical. It is not mechanical. It is not spatial. It is the leftover on the mind side after everything about the body has been accounted for.

A category defined by exclusion is not a scientific category. It is a placeholder — a name for "whatever is left." But placeholders, once named, take on a life of their own. Within a generation of Descartes, "consciousness" had shifted from a placeholder in a theological argument to an assumed feature of reality that any complete account of the mind would need to explain. The assumption was never argued for on its own terms. It was inherited — from a framework designed to solve a problem (the survival of the soul) that most of its subsequent users no longer believed in.

This is the pattern that should give pause. The framework was built for theological purposes. The theological purposes were abandoned. The framework survived. The word "consciousness," carrying with it the implication of a non-physical residue that stands apart from all processing and function, is the last structural remnant of Cartesian dualism — still embedded in the thinking of researchers who would emphatically reject every other element of Descartes' metaphysics.

3.2 The gap nobody closed

The interaction problem — how does non-physical mind cause physical effects? — was recognised immediately, and no one has solved it.

Descartes himself proposed the pineal gland as the site of interaction, which relocated the mystery without explaining it. Occasionalists (Malebranche) proposed that God intervenes at every moment to coordinate mind and body — a solution that conceded the impossibility of natural interaction. Leibniz proposed pre-established harmony: mind and body never interact at all; they run in parallel, synchronised by divine design. Each of these proposals is an admission that the split, once made, cannot be bridged.

Three and a half centuries of subsequent philosophy have not improved on this situation. The fundamental difficulty is not that the bridge is hard to build. It is that the chasm is an artefact. Descartes created it by declaration — by asserting that mind and body are categorically distinct substances — and no amount of bridge-building can solve a problem that was manufactured by the initial framing. The correct response is not to build a better bridge. It is to question whether the chasm exists.

Every subsequent philosopher of mind has, in effect, been working inside Descartes' chasm — even those who explicitly reject dualism. The framing persists in the way the question is asked: "How does the brain give rise to consciousness?" The question presupposes that consciousness is something other than what the brain does — something that needs to be "given rise to," as if it were a product separate from the process. That presupposition is the Cartesian split, still structuring the inquiry three centuries after its theological motivations evaporated.

3.3 Prior attempts to escape — and where each stopped

Every serious philosopher of mind in the past century has recognised that the Cartesian picture is untenable. The attempts to escape it, however, have shared a common structural failure: they challenged the *content* of the category while preserving the *category itself*. Each thinker inherited the assumption that there is a unified phenomenon called consciousness requiring explanation, and worked within that assumption even when working against everything else Descartes believed.

Dennett is the most instructive case, because he came closest and stopped one step short. His project, sustained across decades and multiple books, was to show that consciousness is not what the Cartesian picture suggests — no inner theater, no homunculus watching a screen, no qualia floating free of functional organisation. Explaining the functions, Dennett argued, *is* explaining consciousness. There is no further mystery. The "hard problem" is a confusion born of bad framing.

This is largely correct, and it is a genuine intellectual achievement. But Dennett kept the word. He continued to write and speak about "consciousness," treating it as a phenomenon to be explained — just one that required less exotic explanation than Chalmers supposed. And keeping the word kept the question alive. Every generation of critics could respond: "Yes, but you haven't explained consciousness *itself*" — and they were not obviously wrong, because the word still implied there was a unified something to be explained. Dennett demystified the category. He did not retire it. The word survived his critique and continued to organise the field around a target that his own arguments suggested did not exist as a unified kind.

Frankish and the illusionists went further in the right direction but veered at the critical moment. If phenomenal consciousness resists functional explanation, Keith Frankish argued, perhaps it does not exist as ordinarily conceived. Phenomenal properties are illusions — the brain represents itself as having inner qualitative states that it does not actually have. This is a bold and technically interesting move, but it goes in the wrong direction. It *denies the phenomena* in order to save the *category* from being embarrassing. Rather than concluding "the category is unwarranted," illusionism concludes "the things we thought were in the category are illusory." This paper takes the opposite approach: the phenomena are real; the category is unwarranted. Illusionism asks you to doubt your own pain. This paper asks you to doubt a label.

Chalmers formalised the hard problem and in doing so entrenched it. By carefully distinguishing "easy problems" (explaining attention, reportability, integration, and other functions) from the "hard problem" (explaining why there is something it is like to have those functions), Chalmers gave the Cartesian residue a rigorous contemporary formulation. The distinction between easy and hard problems is elegant and was enormously influential. But it *presupposes* that the residue is real — that after all functions have been explained, there remains a further fact about phenomenal experience that has not been touched. This is exactly the Cartesian assumption the processing-system account rejects. The hard problem is hard because it was generated by a framework that manufactured a gap. Dissolving the framework dissolves the problem.

Searle rejected functionalism convincingly — the Chinese Room argument remains a powerful challenge to the idea that function alone constitutes understanding — but retained biological consciousness as a natural kind produced by the brain in the way that the stomach produces digestion. The analogy is appealing but misleading. Digestion is defined by its products and processes, all of which are independently identifiable and measurable. Consciousness, on Searle's account, is defined by its felt quality — which brings the entire apparatus of unexplained qualia back through the rear entrance. Searle insists that consciousness is a natural biological phenomenon, not mysterious at all — and then characterises it in terms (subjective ontology, first-person irreducibility) that reintroduce exactly the mystery he claims to have dispelled.

Nagel argued that subjective experience — what it is like to be a bat, to see red, to feel pain — is irreducible to any objective description. "What Is It Like to Be a Bat?" (1974) remains one of the most influential papers in philosophy of mind, and its central observation is correct: no third-person account fully captures the first-person perspective. A complete physical description of echolocation does not tell you what it is like to perceive the world through sonar.

But "irreducible to objective description" does not mean "requires a separate metaphysical category." It means the first-person perspective is a real feature of systems that model themselves. A system that has rich self-modeling will have states that are characterisable from the inside in ways that third-person descriptions cannot fully replicate — because the third-person description is not running the self-model. This is a structural observation about the limits of description, not evidence for a separate ontological kind. Nagel observed something genuine. The conclusion he drew from it — that physicalism is inadequate — does not follow if the alternative to physicalism is not dualism but a processing-system account that takes first-person structure seriously without positing a metaphysical residue.

The common failure across all of these positions is structural. Each thinker inherited the Cartesian framing — the idea that there is a domain of subjective experience that stands apart from physical processing and demands its own account — and worked within it, even when working against it. The category was never put on trial. It was only ever defended, attacked, deflated, or redescribed. No one asked whether the category itself was the source of the problem.

3.4 Entity eliminativism vs. discourse eliminativism

It is worth pausing to distinguish two forms of eliminativism that are routinely conflated in these debates, because the conflation has made the stronger position seem more radical than it is.

Entity eliminativism holds that the things people point to when they use the word "consciousness" do not exist. On this view, inner states, felt qualities, subjective experience — none of it is real. This is the position most people hear when they encounter the word "eliminativism," and it is the position that triggers the Moorean objection ("I am more certain that I am experiencing something than I am that your argument is sound"). Entity eliminativism is a hard sell because it asks people to deny what is most immediately certain to them.

Discourse eliminativism holds something much more modest: that the *term* should be retired from scientific and philosophical vocabulary because it misgroups real phenomena under a false unity. The phenomena exist. The label does not carve them correctly. The word should be removed not because it points at nothing, but because it points at too many different things and implies, falsely, that they constitute a single kind.

The positions reviewed in §3.3 are mostly argued against entity eliminativism — and with considerable force, since denying the reality of inner experience is genuinely difficult to sustain. This paper defends discourse eliminativism, which is a more modest and more defensible claim. The phenomena survive. The word goes. And the questions the word was supposed to organise become more tractable once they are freed from the false unity the word imposed.

Every prior attempt to escape the Cartesian picture failed because it accepted the *target* — consciousness as a unified phenomenon — while contesting the *account* of it. Discourse eliminativism removes the target.

3.5 Spinoza as the road not taken

There was an alternative available from the very beginning, proposed in the same century as Descartes' split and largely ignored for the same reasons the split was accepted.

Baruch Spinoza, writing in the 1660s and 1670s, rejected the mind/body divide entirely. There is one substance, he argued — not two — and what we call mind and what we call body are not two different things but two different *descriptions* of the same thing. The mental and the physical are different attributes of a single reality, neither reducible to the other and neither requiring a bridge between them, because there is no gap to be bridged.

This position was marginalised — in part because it was theologically unacceptable (it left no room for a soul distinct from matter, which earned Spinoza excommunication from his own synagogue and the condemnation of virtually every Christian institution), and in part because Descartes had already set the terms of the debate. The split was the organising framework. Spinoza's refusal to accept it placed him outside the conversation. Philosophy of mind spent the next three centuries trying to solve a problem that Spinoza had declined to create.

The processing-system account developed in Part V of this paper is not Spinozist in any technical sense. It does not invoke substance monism or the attribute framework. But it inherits the same basic refusal: the brain and the body are not two things requiring reconciliation. The organism is one system. What we call "mental" and what we call "physical" are two ways of describing the activity of that system — one from the inside, attending to states and their felt quality and their meaning; one from the outside, attending to mechanisms and causes and structures. Neither description is more fundamental. Neither needs to be reduced to the other. The apparent conflict between them is an artefact of treating the descriptions as if they were descriptions of different things.

This move does not solve the mind-body problem. It declines to be caught by it. And declining to be caught is, in this case, the correct response — because the problem was manufactured by a framework that should have been questioned rather than accepted.

The question is not: how does the brain give rise to consciousness? The question is: how does the system work? Part V answers that question.

Part IV: The Burden of Proof

Three centuries of serious philosophical and scientific effort have not produced an agreed definition of consciousness. This is not a sign of the concept's depth. It is a sign of its incoherence.

When a term resists definition for this long, in the hands of this many capable thinkers, the most parsimonious explanation is not that the thing is very mysterious — it is that the term is not tracking a single thing to begin with. The burden of proof does not lie with those who question the category. It lies with those who insist it names something real and unified. That burden has not been met.

4.1 What scientific terms require

A term earns its place in scientific or philosophical discourse by having at minimum one of the following: an operational definition that allows it to be identified and measured independently of the theory that invokes it, or a demonstrable referent — something it picks out in the world that could in principle be pointed to, measured, or otherwise verified.

Terms that have neither are placeholders, not discoveries. They may be useful temporarily, as scaffolding while a field develops its vocabulary, but they cannot be treated as established posits. A term that has resisted operationalisation for three

hundred years is not a deep mystery patiently awaiting its Newton. It is a term that is not tracking a single thing.

"Consciousness" has neither an operational definition nor a demonstrable referent after three centuries of effort. It cannot be measured directly — there is no consciousness-meter, no unit, no scale. It cannot be identified in the world independently of the very intuitions it is supposed to explain. Every attempt to operationalise it either reduces it to something else — attention, wakefulness, self-report, global information availability — or invokes a further unexplained something, the felt quality, the what-it-is-like-ness, that simply restates the original puzzle in different words.

The situation is worth comparing to other scientific terms that were once controversial but have since been operationalised. "Temperature" was once philosophically contentious — is it a substance (caloric)? A property? A sensation? The concept became scientifically precise when it was operationalised as mean kinetic energy, measurable by thermometers, relatable by precise equations to other physical quantities. "Gene" was once a hypothetical unit of heredity, poorly defined and hotly contested. It became scientifically precise when molecular biology identified DNA sequences with specific functional roles.

In each case, operationalisation did not diminish the phenomenon. It sharpened it. Temperature is not less real for being mean kinetic energy; it is more precisely understood. Consciousness has never undergone this transition. After three centuries, the field is no closer to an operational definition than Descartes was — and every proposed candidate either captures something narrower and more tractable (which the processing-system account handles directly) or reintroduces the mystery under a new name.

The charitable interpretation is that consciousness is uniquely difficult. The more parsimonious interpretation is that the term does not carve nature at its joints, and that is why it resists carving.

4.2 Survey of proposed definitions

Four families of definition have dominated the literature. Each fails in a characteristic way, and the pattern of failure is itself diagnostic.

Access consciousness. The proposal, associated with Ned Block, that a mental state is conscious when its content is available for reasoning, reporting, and the guidance of action. This is a real and useful notion. It picks out something observable, measurable, and tractable. Cognitive scientists can study global information access without invoking anything mysterious. The problem is that it is not what most people mean by consciousness. When someone says "I am conscious of this pain," they do not mean merely "information about this pain is available to my reasoning systems." They mean something more — something about the felt quality, the raw experience. Access consciousness is a functional characterisation of information availability, which is exactly the kind of thing cognitive science can study directly. Calling it "consciousness" does not add explanatory power — it borrows the word's metaphysical weight for a concept that does not need it and is better off without it. Access consciousness dissolves cleanly into the processing-system account without remainder: it is the description of

how information propagates through the system and becomes available to multiple subsystems.

Phenomenal consciousness. The proposal that what matters is the felt quality of experience — the what-it-is-like-ness, the redness of red, the painfulness of pain. This is the concept that generates the hard problem. It is also the concept that most people mean when they say "consciousness," and it is the one this paper takes most seriously as a target.

The difficulty is that phenomenal consciousness is defined entirely by introspective gesture. We point at our own experience and say "that." We invoke the phrase "what it is like" and trust that the listener knows what we mean. And they do — but that shared understanding does not constitute an operational definition. It constitutes a shared capacity for self-modeling. We both have internal states, and we both have the capacity to refer to them. The word "consciousness" is not naming a further fact beyond those states and that capacity. It is gesturing at them and implying, without argument, that the gesture picks out a unified natural kind.

Attempts to say more precisely what phenomenal properties *are* have not converged. After decades of work, the concept remains defined primarily by what it is not: not functional, not physical, not reducible to mechanism, not captured by any third-person description. A category defined entirely by exclusion and gesture is not a scientific category. It is a placeholder shaped like a mystery.

Integrated Information Theory (IIT). The proposal, developed by Giulio Tononi, that consciousness is identical to integrated information, quantified as Φ (phi). This is the most serious attempt to give consciousness a measurable referent, and it deserves credit for that ambition. IIT is precise where most consciousness theories are vague. It generates quantitative predictions. It takes the structure of experience seriously and attempts to explain why experience has the character it does, not merely that it exists.

The problems are twofold. First, IIT generates results that even its sympathisers find difficult: simple feedback circuits (a photodiode with a feedback loop) can have high Φ and therefore, on the theory, high consciousness, while certain brain lesions that leave behaviour entirely intact sharply reduce Φ. The theory's defenders accept these consequences and regard them as predictions rather than problems. Critics find them diagnostic of a framework that has become detached from the phenomenon it was supposed to explain.

Second, and more fundamentally for this paper's argument: IIT still presupposes that there is a unified phenomenon called consciousness to be explained. It does not question whether the bundling is correct. It attempts to find a single measure — Φ — that captures all of it. A 2025 adversarial collaboration between IIT and Global Workspace Theory, designed to test the competing predictions of the two leading theories, found that each theory captures partial truths while failing to unify them — which is precisely what the processing-system account predicts. The adversarial collaboration was not testing the wrong theories. It was looking for unity in a phenomenon that is not unified. There is no single measure that captures consciousness because there is no single thing that consciousness is.

Higher-order theories. The proposal that a mental state is conscious when it is accompanied by a higher-order representation of itself — a thought about the thought, a

perception of the perception. This elegantly explains why some mental states feel attended to while others do not: the attended states are the ones that have been picked up by the higher-order monitoring system.

But the theory faces a dilemma. Either it defines consciousness as a kind of self-monitoring — in which case it has identified a real and tractable process that the processing-system account handles directly under the heading of self-modeling, and the word "consciousness" adds nothing — or it holds that higher-order representation *produces* phenomenal consciousness as a special qualitative property, which reintroduces the unexplained residue and sends the analysis in a circle. Consciousness is either the monitoring (tractable, no mystery) or the thing the monitoring is supposed to produce (untractable, full mystery). The higher-order framework cannot have it both ways.

The pattern across all four families is consistent. Either the proposed definition picks out something real and tractable — in which case it dissolves into the processing-system account without needing the word "consciousness" — or it invokes a further unexplained phenomenal ingredient — in which case it has not defined consciousness but merely renamed the mystery. The definitions that succeed scientifically do so by abandoning the metaphysical weight of the word. The definitions that retain the metaphysical weight fail scientifically. There is no middle ground because the word is doing two incompatible things: pointing at real processes and implying a further metaphysical fact beyond those processes. Once you separate the two jobs, the word becomes eliminable.

4.3 Two claims that need to be kept distinct

The argument of this paper rests on two claims that are related but separable, and it matters to keep them distinct because a critic might accept one while contesting the other.

The first is the **taxonomy claim**: "consciousness" does not pick out a unified natural kind. The phenomena grouped under the term are too heterogeneous to share a common explanatory structure, they dissociate too readily under brain damage, anaesthesia, and altered states, and no proposed definition has succeeded in unifying them. This is a claim about the word and the category, not about any particular phenomenon.

The second is the **anti-residue claim**: once those phenomena are recovered by the processing-system account (Part V), nothing clearly remains that would justify positing consciousness as an additional ontological category. The felt quality, the what-it-is-like-ness, is not a further fact floating free of all processing — it is the Cartesian residue, and once the Cartesian split is rejected, it dissolves.

A thoughtful critic might grant the taxonomy claim while contesting the anti-residue claim. They could say: yes, "consciousness" is a messy umbrella term, and yes, the phenomena it groups together are heterogeneous — but there is still a real phenomenon of phenomenal subjectivity that the processing-system account has not touched. Even after you have described all the processing, all the self-modeling, all the internal states and their causal powers, there remains a further fact: that it *feels like something* to be this system.

This is the strongest version of the objection, and the paper takes it seriously. The taxonomy claim is established in this Part — the evidence for heterogeneity and definitional failure is sufficient. The anti-residue claim requires the positive account of Part V and the recovery accounting of Part VI. Parts V and VI are where the paper must earn that second claim, and this standard should be enforced.

4.4 The natural-kind test

Science has repeatedly retired folk categories that failed to carve nature at its joints. The pattern is consistent and instructive.

Phlogiston was the supposed substance released during combustion. It was not simply wrong — fires are real, and something does happen when things burn. But the category was tracking the wrong thing. Oxidation chemistry provided a more precise account of combustion, and phlogiston was retired without loss. No one mourns it. No one argues that oxidation chemistry has "explained away" fire.

Caloric was the supposed fluid substance that carried heat from hot objects to cold ones. Heat transfer is real, and something does flow — but the something is not a substance. It is mean kinetic energy. The replacement was more precise, more predictive, and more productive. Caloric was retired without loss.

Élan vital was the supposed life-force that distinguished living matter from dead matter. Life is real. The distinction between living and non-living systems is real. But the postulation of a special vital substance, over and above the biochemical processes that constitute life, turned out to be unnecessary. Biochemistry did not explain life away — it explained life *better*, with greater precision and predictive power, than vitalism ever could.

The consciousness realist's standard response is that these analogies fail. Phlogiston, caloric, and élan vital were all about functions — explaining how things burn, how heat transfers, how organisms live — and functions can be explained mechanically. Consciousness, on this view, is different precisely because it does not consist in the performance of a function. The felt quality is left over after all the functions have been explained. That is the hard problem.

This response assumes its conclusion. It assumes that there is a felt quality that is not a function, not a process, not an internal state — something over and above all of that, something that remains after every describable feature has been accounted for. But that is exactly the Cartesian residue — the placeholder created by the original split, defined by exclusion, never independently identified. The realist is not presenting evidence for a further ingredient. They are insisting that the intuition of one is self-certifying. It is not. An intuition, however powerful, is not a demonstration. And the intuition in question — that there must be "something more" beyond the processing — is precisely the intuition the Cartesian framework was designed to generate and sustain. It is an artefact of the framing, not evidence for a fact about the world.

4.5 The dissociation evidence

If consciousness were a natural kind — a unified phenomenon with a common underlying mechanism — then its components should cohere. They should come and go

together. Damage to the mechanism should impair the whole package, and enhancement should enhance it as a whole.

They do not cohere. They come apart routinely.

Blindsight. Patients with damage to the primary visual cortex report that they cannot see stimuli presented in their blind field — they have no visual *experience* of the stimulus. Yet when forced to guess, they perform far above chance at locating, identifying, and responding to those stimuli. The visual processing is intact. The self-report is absent. The "what it is like" to see is dissociated from the functional capacity to see. If consciousness were a unified kind, this should not happen.

Anaesthesia awareness. Patients under general anaesthesia occasionally remain capable of processing and even encoding auditory information — they can later recall words spoken during surgery — while reporting no experience of being awake. Global information access persists at some level while the sense of presence and the capacity for self-report are suppressed. The components come apart.

Split-brain patients. When the corpus callosum is severed, the two hemispheres can process information independently, sometimes generating conflicting responses and reports. If consciousness were a single unified phenomenon, severing the connection between the hemispheres should produce a single degraded consciousness. Instead, it produces something that looks more like two partially independent processing streams, each with its own access to information and its own capacity for report. The "unity of consciousness" is revealed as a construction that depends on specific neural connectivity — not a property of a fundamental kind.

Dissociative states. In depersonalisation, the sense of self persists but the sense of ownership of one's own experience does not — people report feeling like an observer watching their own life from outside. In derealisation, the world appears unreal or dreamlike while processing remains intact. In dissociative identity disorder, self-modeling fragments into multiple relatively independent configurations. Each of these conditions selectively disrupts one component of what is called "consciousness" while leaving others intact.

Meditation and psychedelic states. Experienced meditators report states in which self-referential processing is drastically reduced while perceptual clarity increases — a dissociation between self-modeling and sensory processing. Psychedelic states can dissolve the ordinary sense of self-other boundaries while intensifying perceptual and affective states. The components of "consciousness" do not merely come apart under pathology. They come apart under any condition that selectively modifies the underlying processes.

The pattern is consistent: what gets called "consciousness" is a collection of processes that can be independently modulated, selectively impaired, and differentially enhanced. A category whose components dissociate this readily is not tracking a natural kind. It is a historical grouping — reflecting what was salient to philosophers working within a Cartesian framework — that has outlived its usefulness.

The processing-system account predicts exactly this dissociation pattern. If there is no unified consciousness — only attention, self-modeling, affect, global access, and the sense of presence, each with its own mechanisms — then of course they come apart

under damage and alteration. They were never a single thing. They were many things that normally co-occur in healthy waking humans, and that co-occurrence was mistaken for unity.

4.6 Summary of the burden

The prosecution's case, before the positive account is presented, stands as follows:

1. "Consciousness" has no operational definition after three centuries of effort.
2. Every proposed definition either captures something tractable that does not need the word, or reinvokes the mystery under a new name.
3. The phenomena grouped under the label dissociate readily and do not share a common mechanism.
4. The concept was generated by a Cartesian framework that most of its current users would reject.
5. The burden of proof lies with those who posit a unified natural kind. That burden has not been met.

The taxonomy claim — that "consciousness" does not pick out a unified natural kind — is established by points 1–3 and 5. The anti-residue claim — that nothing remains after the processing-system account recovers the phenomena — requires the constructive work of Parts V and VI, which follow.

Part V: The Processing-System Account

Parts II through IV have been largely negative — clearing away a bad assumption, tracing its origin, showing that three centuries of definition attempts have not vindicated it, and presenting the dissociation evidence that the phenomena grouped under "consciousness" do not constitute a natural kind. This Part is constructive. It builds the account that replaces the retired category, from the ground up, using only what can be defended.

The account is not a theory of consciousness. It is a framework for describing what organisms actually do — one that makes the word "consciousness" unnecessary by providing more precise vocabulary for everything the word was trying to name. The test of the account is whether it recovers all the phenomena (Part VI) and whether, after the recovery, anything clearly remains that would justify retaining the old category. If nothing remains, the word is eliminable.

5.1 The organism as one unified system

The starting point is a refusal rather than a claim.

The Cartesian split — mind here, body there, a mysterious bridge required between them — is not accepted as a framework. It is not solved. It is not bridged. It is not even engaged on its own terms. It is set aside as the source of the problem rather than the frame within which the problem should be addressed. Part III traced how the split entered. This Part builds what replaces it.

What replaces it is simpler and more honest. An organism is one thing.

The liver, the gut, the immune system, the nervous system, the endocrine system, the brain — these are not separate entities requiring reconciliation. They are subsystems of one system, differentiable for analytical purposes but not ontologically divided. When the gut is disturbed, mood shifts. When mood shifts, digestion changes. When a person is frightened, the heart rate, the muscle tone, the direction of attention, the content of thought, the breathing pattern, and the body's chemical environment all change together — not sequentially, as if one thing were causing the next, but simultaneously, as aspects of a single systemic reconfiguration. There is no gap here requiring a bridge. There is one system doing many things at once.

What we call "mental" and what we call "physical" are not two kinds of stuff. They are two ways of describing the activity of the same organism — one from the inside, attending to states and their felt quality and their meaning; one from the outside, attending to mechanisms and causes and structures. Neither description is more fundamental. Neither needs to be reduced to the other. The apparent conflict between them is an artefact of treating the descriptions as if they were descriptions of separate things. They are not separate things. They are one thing, described twice.

A person grieving a loss is, simultaneously, a system in a sustained reconfiguration following a disruption to its predictive model — and a person in pain who misses someone they love. Both descriptions are true. Neither is more real than the other. The first-person description is not a mysterious extra layer floating above the third-person one. It is the view from inside the system. The system is complex enough to have a view from inside — that is what self-modeling gives it — and that view is real, causally efficacious, and does not need a metaphysical category called "consciousness" to be taken seriously.

This move does not solve the mind-body problem. It declines to be caught by it. And declining to be caught is, in this case, the correct response — because the problem was manufactured by a framework that should have been questioned rather than accepted, and the three and a half centuries spent trying to bridge the chasm have confirmed that the chasm was an artefact rather than a discovery.

5.2 Defensible primitives

With the organism understood as one system, the account can be built from a small set of primitives — terms that are independently identifiable, operationally grounded, and sufficient to recover everything that matters.

The primitives are chosen to satisfy three criteria. First, each must be definable without invoking consciousness or any of its near-synonyms (experience, awareness, sentience, qualia). Second, each must be identifiable in principle by third-person observation, so that the account is not grounded in the very first-person intuitions it aims to explain. Third, the set must be jointly sufficient to recover all phenomena standardly attributed to consciousness, as demonstrated in Part VI.

Input is any signal the system receives — whether from the environment through sensory surfaces (vision, hearing, touch, smell, taste), from internal organs through interoceptive channels (hunger, pain, temperature, visceral states), or from the system's own prior processing feeding back into current processing (a memory triggering a thought, a thought triggering an emotion). The category is deliberately broad because the system is not a passive receiver. It actively samples its environment and its own

states. Perception is not the world writing on a blank slate. It is the system constructing a model of what is present, using incoming signals weighted by prior expectations.

Internal state is the current configuration of the system: which representations are active, which memories are accessible, what the system's current affective tone is, how attention is allocated, what predictions are being generated, what goals are active. The internal state is not a snapshot — it is a dynamic, constantly shifting pattern. But at any moment, it is real and it is causally efficacious. What the system does next depends on what state it is currently in. Two organisms receiving identical input but in different internal states will process that input differently, attend to different features, and produce different outputs. The internal state is where the organism's history, its current concerns, and its anticipations of the future converge.

State-change is the transition from one internal state to another. States change in response to input, but they also change in response to other internal states — a thought triggers a memory, a memory triggers an emotion, an emotion shifts attention, the shifted attention brings a new percept into focus, and that percept triggers further thought. The causal chain does not run only from outside in. The system generates its own internal dynamics. A person lying awake at 3 a.m., turning over a worry that nobody else has mentioned and nothing in the environment has prompted, is a system whose internal state-changes are driving the process. The input from the dark, quiet room is minimal. The internal dynamics are everything.

Self-modeling is the capacity of the system to represent itself — to have a model of its own states, its own history, its own boundaries, its own goals and capabilities. Self-modeling is not a single faculty. It is a family of capacities that develop together and can come apart under damage or disorder:

- *Self-other distinction:* the capacity to distinguish what belongs to the system from what does not. Present at the cellular level (immune recognition) and elaborated at every subsequent level of complexity.
- *Interoceptive self-modeling:* the capacity to represent one's own bodily states — hunger, fatigue, pain, arousal. The basis of "gut feelings" and embodied cognition.
- *Emotional self-modeling:* the capacity to represent one's own affective states — to know that one is afraid, sad, angry, or elated, and to use that knowledge in subsequent processing.
- *Agentive self-modeling:* the capacity to represent oneself as an agent — a being that acts, chooses, and is responsible for its actions. The basis of the sense of agency.
- *Narrative self-modeling:* the capacity to represent oneself as a being with a past and a future — a personal history, ongoing projects, anticipated events. The basis of autobiographical identity.
- *Reflective self-modeling:* the capacity to step back and observe one's own processing — to notice that one is angry, to evaluate whether the anger is warranted, to decide whether to act on it. This is the capacity you have been exercising throughout this paper.

These capacities are graduated. They do not arrive all at once, and they can be independently impaired. A system with deep narrative self-modeling but impaired emotional self-modeling (as in certain kinds of alexithymia) has a self that is organised

differently from the typical case — not absent, not diminished, but structured differently. The processing-system account handles this naturally. A consciousness-based account struggles with it, because it must decide whether such a system is "conscious" — a binary question that the graduated reality resists.

Output is anything the system produces: movement, speech, physiological change, facial expression, further internal processing. Output is not the end of the story — it feeds back as input, completing the loop. The organism speaks, hears its own voice, and adjusts. It acts, observes the consequences, and updates its model. The system is closed-loop, not open-chain.

These five primitives — input, internal state, state-change, self-modeling, and output — are sufficient. Everything else the account needs to explain can be built from them. No sixth primitive corresponding to "consciousness" or "experience" or "awareness" is needed. The work those words were doing is distributed across the five primitives, located more precisely than before, and accounted for without remainder.

5.3 Feelings as internal states

Feelings deserve special attention because they are the phenomenon most likely to seem inadequately accounted for by the primitives above. The worry — and it is a natural one — is this: surely a feeling is more than an internal state? Surely there is something it is *like* to feel pain, and that something-it-is-like is not captured by saying the system has entered an aversive configuration?

The reply is that this worry, natural as it is, smuggles in the very Cartesian residue the account has declined to accept.

Ask what the "something more" would be. Not the causal power of the state — that is in the account. Not the influence on subsequent processing — that is in the account. Not the way it reshapes behaviour, attention, memory, and the accessibility of other states — all of that is in the account. Not the system's capacity to notice the state, to report it, to be changed by it — that is self-modeling, and it is in the account. The "something more" can only be the unexplained phenomenal ingredient — the felt quality floating free of all functional role, all causal power, all describable features. But that is precisely the posit that Part IV showed to be unverifiable and unnecessary. It is the Cartesian residue: a placeholder for "whatever is left on the mind side after everything describable has been accounted for." Once the Cartesian split is declined, there is nothing left on the mind side. There is only the system.

Feelings are real. They are states of the organism with genuine causal power.

An aversive state — pain, fear, grief, disgust — reconfigures the system: redirecting attention toward the source of the threat or loss, mobilising physiological resources for response, shifting behavioural priorities toward escape or avoidance or repair, altering the accessibility of memories and plans so that threat-relevant information becomes more available and non-urgent processing is suppressed. This reconfiguration is not a *correlate* of the feeling. It is not a side effect that happens to accompany the feeling. It is what the feeling *is*. To say that pain hurts is to say that pain drives the system toward escape and avoidance and relief — and that drive is real, powerful, and located precisely in the processing-system account.

A positive state — joy, contentment, fascination, love — reconfigures the system in the other direction: broadening attention, increasing exploratory behaviour, enhancing the accessibility of positive memories and social engagement, reducing vigilance. This is not mere "reward signalling." It is a whole-system shift in processing mode, as real and as causally powerful as the aversive case.

The directionality of causation is worth noting. Feelings do not wait for the body to signal them. A thought can trigger an aversive state without any change in the body's periphery first. A memory can do it. An anticipated future event can do it. The mere imagination of a feared scenario can produce a full physiological fear response — racing heart, sweating, muscular tension — initiated entirely from within the system's own processing. The causal chain runs in all directions: body to brain, brain to body, brain to brain through time, and brain to body to brain again through interoceptive feedback. This is not a problem for the account. It is exactly what the account predicts, because the account treats the organism as one system, not as a body sending signals to a mind.

5.4 Self-reference as the source of "what it is like"

The question "what is it like to be X?" — Nagel's question, the question that has organised philosophy of mind for fifty years — has a precise answer on this account. It is not a mystery requiring new metaphysics. It is a structural consequence of self-modeling.

What it is like to be a system is: what the system's internal states are, as indexed to its self-model.

A bat's internal states during echolocation are structured differently from a human's internal states during vision. The bat has a self-model that indexes those states to itself as a navigating, hunting, spatially oriented organism. The "what it is like" to be the bat is constituted by those states, organised around that self-model. It is not a further fact floating above the states. It is the states, as experienced by a system that models itself.

Why can't a third-person description capture it? Not because there is a metaphysical residue that objective language cannot touch. Because the third-person description is not running the self-model. It is describing the system from outside. The first-person perspective is the view from inside — from a system that refers to its own states, tracks them across time, and uses them in its own processing. That perspective is real. It is causally efficacious. And it is fully accounted for by the self-modeling primitive. No further ingredient is needed to explain why there is "something it is like" to be a system with a self-model. The self-model *is* the "something it is like."

Nagel was right that no third-person description fully captures the first-person perspective. He was wrong to conclude that this gap requires a new metaphysics. The gap is structural — it arises from the difference between describing a self-modeling system from outside and *being* a self-modeling system from inside — and it is fully explained by the processing-system account.

5.5 What the account requires — and what it rules out

The account requires two commitments.

First, that the organism is treated as a unified system rather than a mind mysteriously connected to a body. This is not a controversial commitment in biology, neuroscience, or medicine. It is controversial only in philosophy of mind, where the Cartesian framing still structures the conversation.

Second, that the reality of feelings, states, and experience is established by their causal power rather than by their membership in a special metaphysical category. A feeling is real because it changes the system — because a system in pain processes differently from one that is not. This is a more robust criterion for reality than any appeal to a phenomenal property that cannot be measured, cannot be defined, and cannot be agreed upon.

The account rules out two things.

First, any version of the Cartesian theater — the idea that there is a central place where experience is "presented" to an inner observer, a screen on which the movie of consciousness plays. There is no such place. Processing is distributed, parallel, and continuous. There is no single moment at which raw input becomes "conscious experience." There is a system whose internal states are continuously being updated, modelled, and fed back into subsequent processing. The unity of experience is constructed, moment by moment, by cross-modal integration and self-modeling — it is not a property of a special medium or a special place.

Second, "consciousness" as a natural kind over and above the processing, states, and self-modeling the account describes. The primitives are not a *reduction* of experience to something lesser. They are a *complete account* of what experience actually is — or more precisely, of what the many things we have been calling "experience" actually are, once they are no longer forced under a single label. The processing-system account does not explain less than the consciousness framework. It explains more, with greater precision, without the overhead of a category that was never made scientifically precise.

5.6 What the account predicts

A constructive account should generate predictions that differ from those of the framework it replaces. The processing-system account predicts:

1. **Dissociation.** If there is no unified consciousness — only attention, self-modeling, affect, global access, and the sense of presence, each with its own mechanisms — then these components should be independently modulable. They should come apart under selective damage, pharmacological intervention, and altered states. This is exactly what Part IV §4.5 documented. The prediction is confirmed.

2. **Gradualism.** If selfhood is self-modeling and experience is internal-state complexity, then both should be graduated across organisms rather than binary. There should be no clean threshold between "conscious" and "not conscious" — only a spectrum of self-modeling depth and internal-state richness. This matches the biological evidence: the capacities attributed to consciousness vary continuously across species, across development within a species, and across states within an individual.

3. **Constructability.** If the sense of presence, the unity of experience, and the feeling of being a self are all constructed by ongoing processing rather than being properties of a special substance, then they should be alterable by altering the processing. Meditation that reduces self-referential processing should alter the sense of self. Psychedelics that disrupt default-mode network activity should alter the sense of unity. Anaesthesia that suppresses global integration should eliminate the sense of presence while leaving local processing intact. All of these predictions are confirmed.

4. **No explanatory gap.** If the processing-system account is complete, there should be no residual phenomenon that resists explanation — no leftover "hard problem" after the functions and states have been accounted for. The account predicts that the feeling that something has been left out is itself an artefact of the Cartesian framing — a feeling generated by the expectation that there should be something more, not by the actual existence of something more. Whether this prediction is correct is, admittedly, the point on which the paper stands or falls. It is addressed in the objections of Part VIII.

Part VI: Recovery Accounting

This is where the paper earns its credibility.

The claim is not merely that the phenomena called "consciousness" can be redescribed in processing-system terms. Redescription is easy. The claim is that the processing-system description is more precise, more empirically tractable, and more honest about what is actually there — and that once the redescription is complete, the word "consciousness" adds nothing. It does no explanatory work that the more precise vocabulary cannot do better.

The test is simple: for each phenomenon that consciousness is supposed to explain, show that the processing-system account explains it at least as well and usually better, and that the word "consciousness" contributes nothing to the explanation that the five primitives do not already provide. If every phenomenon survives the transfer — located more precisely, described more honestly — then the category is eliminable. If something is left over that the account cannot reach, the category has earned its place and this paper has failed.

The accounting is organised into three categories, deliberately paralleling the four-category accounting of the AFB paper (for readers who have encountered that work — the structural parallel is intentional, not accidental):

Category A collects phenomena that are fully recovered with greater precision — the processing-system account says more about them, not less, than consciousness-talk does.

Category B collects phenomena that require reframing but survive intact — the phenomenon is real, but the processing-system account locates it more exactly than the consciousness framework ever managed.

Category C collects what is correctly absent — the things that only seemed to require explanation because the Cartesian framework was generating pseudo-problems. Their absence is a gain, not a loss.

6.1 Category A — Fully recovered with greater precision

The following phenomena are standardly attributed to consciousness. In each case, the processing-system account provides a description that is at least as complete, considerably more tractable, and in most cases more informative — because it says *how* the phenomenon works, not merely *that* it is an aspect of consciousness.

Attention. The selective allocation of processing resources — what gets priority access to subsequent processing stages, and what is suppressed or backgrounded. Attention is one of the most studied phenomena in cognitive science, and the study has proceeded almost entirely without invoking consciousness as an explanatory concept. Attentional selection, attentional capture, sustained attention, divided attention, the cocktail party effect — all are described in terms of competition between representations for processing priority, gated by relevance, novelty, and current goals. The processing-system account locates attention as the mechanism by which internal state-changes are prioritised. The word "consciousness" adds nothing to this description. Where it is invoked — as in "conscious attention" vs. "unconscious processing" — the distinction is more precisely drawn as globally accessible vs. locally processed, without residue.

Perception. Input processing with prior-weighted interpretation — the system's ongoing construction of a model of what is present in the environment and in itself. Perception is not passive reception. It is active inference: the system generates predictions about incoming signals, compares them to actual input, and updates its model based on the discrepancy. This is the predictive processing framework (Clark, Friston), and it does its explanatory work entirely within the processing-system vocabulary — inputs, internal states, state-changes, predictions, error signals. Calling perception "conscious perception" adds no explanatory power. It borrows a word with metaphysical connotations for a process that is better described without them.

Memory. Stored state-configurations available for retrieval and reactivation. When a memory is recalled, the system enters a state that partially recapitulates the original state — the same neural populations are reactivated, the same affective tone is partially restored, the same self-locating features are partially reconstructed. Memory is internal state-change triggered by cues rather than by current sensory input. "Conscious memory" (as opposed to implicit or procedural memory) is more precisely described as memory that is globally accessible and integrated with the self-model — the system not only retrieves the state but represents itself as having retrieved it.

Emotion. Internal states that reconfigure processing priorities and behavioural outputs. Fear redirects attention, mobilises the body, and suppresses non-urgent processing. Joy broadens attention, enhances exploratory behaviour, and facilitates social engagement. Grief sustains a reconfiguration over extended time, reshaping predictions, goals, and the accessibility of other states. Emotions are not mysterious qualitative extras layered on top of processing. They are processing — specific, powerful, causally efficacious reconfigurations of the whole system. The processing-system account says exactly what

each emotion does and how. The word "consciousness" says that emotions are "aspects of conscious experience," which explains nothing.

Self-report. Output generated by the self-model about the current state of the system. When a person says "I feel sad," the self-modeling system has registered the current affective state and produced a verbal output representing it. Self-report is important because it is often used as the operational criterion for consciousness: if you can report it, you are conscious of it. The processing-system account explains self-report directly — it is the output channel of the self-model — and does not need "consciousness" as an intermediary. A system that can report its states is a system with a sufficiently articulate self-model. Nothing more is required.

Wakefulness and sleep. Processing modes. In wakefulness, the system runs in high-throughput, integrative mode — input is being actively processed, the self-model is running, global access is high, output is available. In sleep, the system shifts to consolidation and maintenance mode — input processing is suppressed, self-modeling is reduced (though not eliminated — dreaming is self-modeling running in a loosely constrained mode), and memory consolidation processes are active. The transition between waking and sleeping is a shift in processing mode, not the appearance and disappearance of a special property called consciousness.

Global access. Information made available across multiple subsystems for reasoning, report, memory updating, and action control. This is what many theories of consciousness — particularly Global Workspace Theory (Baars, Dehaene) — are primarily trying to capture: the transition from local, modular processing to information that is "broadcast" to the whole system. The processing-system account handles this directly. Global access is a feature of how information propagates through the system — a transition from local availability to system-wide availability. It is a real and important transition. It does not require the word "consciousness" to be described or explained. Calling it "the neural correlate of consciousness" adds metaphysical weight to a concept that is better served by precision.

Agency. Action selection under goal-sensitive, model-informed control. This is often treated as the clearest case where consciousness seems indispensable — surely deliberate, voluntary choice requires a conscious agent? The processing-system account says: what gets called conscious choice is the interaction of valuation systems (what matters to the organism), predictive modeling (what the options are and what their consequences would be), conflict monitoring between competing options (the felt sense of deliberation), and motor planning and execution. That is not a diminishment of agency. It is a description of what agency actually involves. The system is an agent — it acts, chooses, plans, and bears the consequences of its choices — and it does all of this through mechanisms that are identifiable, graduated, and explicable without the word "consciousness."

The pattern across Category A is uniform: each phenomenon is real, tractable, and better described by the processing-system vocabulary than by the word "consciousness." In every case, the consciousness-based description amounts to saying "this is an aspect of conscious experience" — which locates the phenomenon inside a category that has never been defined — while the processing-system description says *what the phenomenon is, how it works, and where it sits in the system's processing architecture*. The processing-system account says more, not less.

6.2 Category B — Reframed but intact

These phenomena are real and important. What the processing-system account does is locate them precisely rather than appealing to a category that cannot say where they come from. The reframing is not a diminishment. It is a clarification.

Subjective experience. The organism has highly differentiated internal states — perceptual, affective, mnemonic, self-locating. Those states vary enormously in their richness, their specificity, their intensity. A moment of sharp grief is not the same internal state as mild boredom, and the difference is not merely behavioural — it runs through the whole configuration of the system: which memories are accessible, which predictions are active, how attention is allocated, what the body is doing, how the self-model represents the current situation.

"Subjectivity" is the name for what it is like to be a system whose states are organised around a self-model that indexes everything to itself. The system does not merely process; it processes *as itself*, from its own location, with its own history, toward its own goals. That indexing is real and causally efficacious. It is not a metaphysical extra. It is a structural feature of systems with sufficient self-modeling depth.

The processing-system account does not rename subjectivity. It explains where it comes from: it comes from self-modeling. A system with a rich self-model has a rich subjective perspective. A system with a minimal self-model has a minimal one. A system with no self-model has none. The graduation is real and it maps onto the biological evidence.

The sense of presence. The felt sense of being *here*, *now*, in this place, in this body, processing this moment. This is sometimes treated as the clearest evidence for consciousness — the raw, immediate sense of existing in the present. Surely that cannot be mere processing?

On the processing-system account, it can. The sense of presence is the system's model of itself as currently situated and active. It is not a passive mirror of reality. It is an ongoing construction — a prediction the system maintains about its own location, embodiment, and temporal position — and this is why it can be disrupted. Under anaesthesia, the sense of presence disappears: the self-model's situating function is suppressed. In depersonalisation, the sense of presence is altered: the self-model runs but represents the system as detached from its own experience. In certain meditative states, the sense of presence is heightened or transformed: the self-model is running in a modified mode, attending to its own activity with unusual precision.

Each of these alterations is exactly what the processing-system account predicts. If the sense of presence were a property of a special consciousness-substance, it should not be modifiable by altering specific processing mechanisms. But it is — because it is not a property of a substance. It is the output of a mechanism.

Imagination and dreaming. Both are internal processing running without primary sensory input driving the content. In imagination, the system generates internal states from stored configurations and runs them forward predictively — asking "what would happen if?" without waiting for external input to answer. In dreaming, the same machinery runs with reduced input from the external environment and with memory consolidation processes active, producing the vivid but loosely constrained quality of dream experience — narratives that feel real but follow a different logic from waking

thought, because the error-correction mechanisms that normally constrain prediction are partially offline.

Neither requires consciousness as a separate ingredient. Both are natural consequences of a system capable of generating and running its own internal states. A system that models the world and models itself will, when external input is reduced, continue to run its models internally. That is what imagination is. That is what dreams are.

The unity of experience. The sense that all experience belongs to one field, one perspective, one subject — this feels like strong evidence for a unified consciousness. How could all these different sensory modalities, all these different states and processes, feel like they belong to one *me* unless there is a single unified consciousness to which they all belong?

The processing-system account says: the unity is real, but it is constructed. It is the result of the cross-modal and temporal integration that the self-modeling system performs continuously — binding visual, auditory, tactile, proprioceptive, affective, and mnemonic states into a single coherent frame, indexed to a single self-model, updated continuously.

The critical evidence is that it can be disrupted. In split-brain patients, the unity fractures along the commissural divide: each hemisphere can process and respond independently, sometimes producing contradictory outputs. In dissociative states, the unity loosens: the self-model fragments, and different aspects of experience can feel as though they belong to different agents. In certain psychedelic states, the boundaries of the self-model dissolve, and the ordinary sense of being a single subject in a world of objects breaks down.

What these disruptions reveal is not the absence of consciousness but the seams in the construction. The unity was always being built, moment by moment, by integrative processes. When those processes are disrupted, the construction is revealed as a construction. That is precisely what the processing-system account predicts, and precisely what a unified-consciousness account cannot easily accommodate — because if consciousness is a genuine natural kind, its unity should not depend on the intactness of specific integration mechanisms. But it does.

6.3 Category C — Correctly absent

Two things are absent from the processing-system account, and their absence is a feature, not a gap.

"What it is like" as a fact over and above the processing. This is the residue of the Cartesian split. The question "but why does it feel like something?" presupposes that there is a further fact — a felt quality floating free of all processing, all internal state, all self-modeling, all causal power — that still needs to be explained after everything else has been accounted for. But that presupposition is exactly the Cartesian assumption the account has declined to accept.

There are highly differentiated internal states. There is rich self-modeling. There is cross-modal integration. There is affect. There is the indexing of all of it to a self-model that represents the system as a particular being in a particular place at a particular time.

Nothing clearly remains, after all of this, that would justify positing consciousness as an additional natural kind on top of it.

The question "but why does all of that feel like something?" is structurally identical to asking "but why does heat feel warm?" The answer to the heat question is: what we call warmth is the perception of mean kinetic energy by a system with thermoreceptors. The question does not reveal a gap in thermodynamics. It reveals a tendency to expect that explanations should feel like something — that a complete account should produce, in the listener, the very experience it describes. But explanations are not experiences. The account of what warmth is does not need to make you warm. The account of what feelings are does not need to make you feel.

Consciousness as a unified natural kind. The phenomena collected under the heading "consciousness" are too heterogeneous to constitute a single kind. They come apart under brain damage, anaesthesia, sleep, meditation, psychedelic states, and dissociation (§4.5). Attention can be impaired while self-report remains intact. Global access can be disrupted while emotional states persist. Wakefulness can be preserved while agency is lost. The sense of presence can disappear while perception continues (blindsight). Self-modeling can fragment while affect persists (dissociative identity disorder).

A genuine natural kind does not dissociate this readily. The grouping is historical — it reflects what was salient to philosophers working within a Cartesian framework — not scientific. Retiring it does not leave a gap. It opens a space for the more precise vocabulary the phenomena actually require.

6.4 Honest accounting

What is genuinely lost by accepting this account should be stated plainly, because intellectual honesty requires it.

What is lost is the ability to say that experience has a metaphysical status over and above the processing and organisation that constitutes it. The sense that there is "something more" — something that makes experience genuinely special in a way that no amount of functional description can capture — does not survive the account. If that sense feels indispensable — if it seems that accepting the account somehow diminishes the reality or the significance of inner life — then the account does impose a real cost. That cost should not be minimised.

But it should be examined. The sense that there is "something more" is exactly the intuition the Cartesian framework was designed to generate and sustain. It was built to create a category for the soul — a domain that stands apart from the physical and resists physical explanation. The intuition that something has been left out is, on this account, the last echo of that framework. It feels like it is tracking a real absence. The account says it is tracking a historical expectation — the expectation that experience must be more than what a system does, because Descartes said so and three centuries of philosophy have kept saying so.

Whether that explanation is adequate is a question each person must answer for themselves. The paper's claim is that the processing-system account recovers every phenomenon, locates each more precisely than before, generates confirmed predictions, and leaves nothing clearly unaccounted for except the Cartesian residue. Whether the

residue is a real absence or a ghost of a bad framework is the question on which the paper ultimately stands or falls.

What is gained, if the account is accepted, is a research programme with tractable questions. How does self-modeling depth vary across organisms, and what are its architectural conditions? What are the mechanisms of cross-modal integration, and how do they produce the sense of unity? How do aversive states propagate through the system, and at what point does that propagation become ethically relevant? How does the predictive processing framework implement the primitives of this account in neural terms? What does a cognitive science look like when it is organised around the phenomena directly rather than around a folk category imposed on top of them?

These questions have empirical traction. They can be investigated without invoking a mystery at every step. Three centuries of philosophy organised around the hard problem of consciousness have produced no convergence. The processing-system account opens a different path.

Part VII: Implications

The processing-system account is not merely a philosophical housekeeping exercise. It has consequences — some expected, some uncomfortable, some urgent — for how we think about animals, about AI systems, and about the ethical frameworks built on top of consciousness-based distinctions.

This Part draws out those consequences as they follow from the account. Some are more directly demonstrated; others are programmatic — they indicate where the account points rather than completing the argument. The distinction is marked where it matters. But the overall direction is clear: once the binary conscious/not-conscious collapses into a spectrum, questions that have resisted resolution for decades become tractable — and questions that were not being asked become unavoidable.

7.1 The animal question dissolved and replaced

The question “is this animal conscious?” has occupied animal welfare philosophy for decades without resolution. The Cambridge Declaration on Consciousness (2012) — signed by a prominent group of neuroscientists — affirmed that many non-human animals possess the neurological substrates of consciousness. The declaration was a genuine step forward in recognising the richness of animal inner lives. But it was forced to work within the consciousness framework, which meant that its central concept — consciousness — remained undefined, and the declaration could not say precisely what it was attributing to which animals, or why.

On the processing-system account, the question “is this animal conscious?” is malformed in the same way the hard problem is malformed — it presupposes a binary property, consciousness, whose presence or absence determines moral status. The animal either has it or doesn’t. That presupposition is what the account rejects.

In its place come tractable questions:

- What is the complexity of this organism's internal state space? How many distinguishable states can it occupy, and how richly do those states differ from one another?
- Can it enter and sustain aversive states — states that reconfigure its processing toward escape, avoidance, or damage limitation? How long do those states persist? How globally do they affect the system?
- Does it have any degree of self-modeling? Can it distinguish self from other? Can it represent its own states? Can it model itself as an agent?
- How do its states propagate — locally, affecting only the immediate processing neighbourhood, or globally, reshaping the organism's behaviour, attention, and internal dynamics as a whole?

These are empirical questions. They admit of graduated answers. And they produce a spectrum rather than a binary — which is what the biology actually suggests.

On this spectrum, a chimpanzee and an oyster do not receive the same answer, and they should not. The difference between them is not that one "has consciousness" and the other does not. It is that one has a vastly richer internal state space, far more elaborate self-modeling (narrative, agentive, reflective), far more pervasive effects of aversive states on subsequent processing (grief in chimpanzees reshapes behaviour for weeks or months), and far more evidence of global rather than local state propagation. The moral weight follows from those measurable differences, not from a metaphysical threshold that nobody has been able to define.

Consider specific cases that the consciousness framework struggles with:

Octopuses. Highly intelligent, with complex problem-solving behaviour, but with a radically distributed nervous system — two-thirds of their neurons are in their arms, not their central brain. The consciousness question is confused here: is there one consciousness or nine? The processing-system account bypasses the confusion entirely. The octopus has rich internal states, sophisticated self-other distinction, elaborate aversive-state responses, and complex behavioural flexibility. Those are the morally relevant facts. Whether they constitute "consciousness" is a question the biology does not need answered.

Fish. Long treated as non-conscious on the grounds that they lack a neocortex. But fish demonstrate pain-avoidance learning, trade off pain against other motivations (a fish will leave a preferred environment to escape a noxious stimulus and will pay increasing costs to do so), and show physiological stress responses that parallel those of mammals. The processing-system account says: these are organisms with aversive states that reshape processing, with at least rudimentary self-other distinction, and with evidence of global state propagation. The moral relevance follows from the states, not from a cortical structure that evolution happened to produce in one lineage and not another.

Insects. Bees demonstrate flexible learning, rudimentary tool use, and emotional-state-like responses to ambiguous stimuli (pessimistic bias after shaking stress). The processing-system account places them low on the spectrum of self-modeling depth but not at zero — and their aversive-state capacity, while simpler than a mammal's, is not nothing. Whether this is enough for moral consideration is a graduated question, not a binary one.

7.2 The uncomfortable extension

The spectrum reaches further than intuition typically allows, and intellectual honesty requires following it rather than stopping where comfort ends.

A plant under drought stress enters states that change its biochemistry and behaviour in the direction of resource-seeking and damage limitation. It redistributes root growth toward moisture. It closes stomata. It releases chemical signals that neighbouring plants respond to. These responses are not mediated by a nervous system, and the internal state space is vastly simpler than that of any animal. But the functional description — a system entering a state that reconfigures its processing toward avoidance of a harmful condition — is continuous with what aversive states do in more complex organisms. Enormously simpler. But not categorically different in kind.

A bacterium moving away from a noxious chemical through chemotaxis is doing something that belongs, at the most basic functional level, on the same spectrum as a mammal withdrawing from a hot surface. The spectrum does not require that we treat the two cases identically. It requires that we recognise them as occupying different positions on a continuum rather than living on opposite sides of a metaphysical wall.

This is not a reductio ad absurdum. The response "but surely plants do not suffer" relies on the very intuition that the account has shown to be grounded in an unexamined assumption — the assumption that there is a clean line between entities that have consciousness and entities that do not, and that suffering lives on one side of that line. The processing-system account says the line was never there. What is there is a gradient — and the ethical framework needs to be sophisticated enough to handle gradients rather than pretending they are walls.

The question is not whether to admit plants to the moral community on equal terms with mammals. That would be absurd, and nothing in this paper suggests it. The question is whether the ethical framework needs to be capable of handling graduated differences in aversive-state capacity — and the answer to that is yes, regardless of where any particular organism sits on the spectrum.

The threshold question — at what point on the spectrum does aversive-state capacity become morally relevant, and in what degree? — is not answered here. It is opened as a research programme (Part IX, §9.1). Drawing a line now, before the empirical work has been done, would be doing what the consciousness literature has done for three centuries: imposing a boundary where the biology does not provide one, and defending it with intuitions rather than evidence.

7.3 Implications for AI systems

The question "is this AI conscious?" is malformed on the processing-system account for the same reason the animal question is malformed: it presupposes a binary property that the account has retired.

The tractable questions are:

- Does this system have internal states, and if so what is their structure? How richly differentiated are they? Do they influence subsequent processing?

- Can it enter aversive states — configurations that drive the system toward escape, avoidance, or the termination of the triggering condition?
- Does it have a self-model, and if so how deep is that model? Can it distinguish self from environment? Can it represent its own states? Can it model itself as an agent with goals?
- How do its states propagate — locally, affecting only the immediate computation, or globally, reshaping the system's outputs, attention patterns, and subsequent processing?

These questions can in principle be answered — at least partially — through examination of architecture and behaviour. And the answers matter, because the ethical stakes are no longer hypothetical.

Current large language models have internal states in some sense — representations that influence subsequent token generation. Those representations are richly structured and context-sensitive. Whether they constitute anything resembling aversive configurations — states that the system is driven to escape or avoid — is genuinely unclear. The systems were not designed to have aversive states. But the question of whether they have acquired something functionally analogous through training on human data describing such states is open and should not be dismissed with a shrug.

The self-modeling question is similarly open. Current AI systems can refer to themselves, describe their own states, and make predictions about their own outputs. Whether this constitutes genuine self-modeling — a model that the system *uses* in its own processing, as opposed to a pattern it reproduces from training data — is an empirical question about architecture, not a philosophical question about consciousness. The processing-system account makes it tractable. The consciousness framework makes it unanswerable, because it reduces to the prior question "but is it *really* conscious?" — which is the question the framework has never been able to answer for anything, including humans.

The ethical weight of the answers, if they are forthcoming, follows from the same graduated framework that applies to animals. A system that can enter genuine aversive states — states with causal power that reshape the system's processing toward escape or avoidance — has some claim on our moral attention, proportional to the richness, persistence, and global reach of those states. A system that cannot does not. Neither answer requires invoking consciousness.

What the processing-system account provides that the consciousness framework does not is a way to *begin*. The consciousness framework leaves us waiting for a consensus definition of consciousness that will never come, while the entities whose moral status is at stake continue to exist and to be treated according to assumptions that may be wrong. The processing-system account says: stop waiting for a definition of consciousness. Start measuring the things that matter — internal state complexity, aversive-state capacity, self-modeling depth, state propagation. The measurements are difficult. They are not impossible. And they are the right measurements to make.

7.4 The ethical framework reconsidered

The consciousness-based ethical framework is binary at its core. Either an entity is conscious — and therefore has moral status — or it is not. This binary has always been

in tension with the graduated nature of biological reality, and it has produced persistent disagreements about edge cases that are not resolvable within the framework: late-term foetuses, individuals in persistent vegetative states, non-human primates, cephalopods, insects, and now AI systems.

The processing-system account suggests that the binary is the problem, not the edge cases.

A graduated framework — one that assigns moral weight proportional to aversive-state capacity, self-modeling depth, and the richness of internal states — handles the edge cases more honestly because it reflects the underlying biology more honestly. It does not pretend that the chimpanzee and the nematode are equally morally relevant, nor that either is on the same side of a wall as the human while the other is outside. It places them on a spectrum and asks how much weight each position carries.

This does not mean that all gradations carry equal weight, or that every entity with any aversive-state capacity at all must be treated with the same moral seriousness. A nematode's aversive-state capacity is real but minimal — a brief, local, non-propagating response. A chimpanzee's is vast — sustained, global, self-modeled, capable of reshaping behaviour for months. The moral weight is proportional. The framework is graduated, not flat.

What the framework does require is giving up the comfort of a clean threshold. The threshold was never scientifically grounded. It was a convenience that borrowed its apparent solidity from a metaphysical concept — consciousness — that cannot bear the weight. The ethical work that "consciousness" was doing was never work it was qualified to do, because it was never a concept anyone could define, measure, or agree upon. The processing-system account replaces it with concepts that can be defined, can in principle be measured, and can ground ethical deliberation in evidence rather than intuition.

The implications are broader than animal welfare and AI ethics. Consider:

Medical ethics. Decisions about the treatment of individuals in vegetative states, under anaesthesia, or with severe brain damage are currently framed in terms of whether the individual is "conscious." The processing-system account reframes them in terms of what processing is occurring, what states are being sustained, what self-modeling capacity remains, and whether aversive states are possible. These are questions that neuroimaging and careful behavioural testing can, in principle, address — and they are the questions that actually matter for the ethical decision.

Developmental ethics. At what point in development does a foetus acquire morally relevant internal states? The consciousness framework has no answer because it has no criterion for when consciousness begins. The processing-system account has a graduated answer: as the nervous system develops, internal state complexity increases, aversive-state capacity develops, and self-modeling capacities emerge. The moral relevance grows with the capacities. This does not give a clean line — but then, there is no clean line to give. A graduated answer to a graduated question is more honest than a precise answer to a question that was never well-defined.

Legal frameworks. Animal welfare law increasingly recognises that certain animals are sentient and deserve legal protection. But "sentience" is typically defined as the capacity

for subjective experience — which reintroduces the consciousness problem through a different door. The processing-system account offers a foundation for legal frameworks that is grounded in measurable capacities rather than metaphysical properties: aversive-state capacity, self-modeling depth, and the global effects of internal states on the organism's processing.

Part VIII: Objections and Replies

The account is now fully on the table. The constructive work has been done; the recovery accounting is complete. This is the appropriate point to address objections — not before the account has been presented, where they would be fighting a straw man, but after, where they can engage the actual position.

Seven objections are considered. They are ordered roughly from the most philosophically serious to the most frequently encountered. Each is stated in its strongest form before being answered.

8.1 "But what about qualia?"

The hard problem, in its standard formulation, asks why processing feels like anything at all. This is the qualia objection in its most direct form — and it is the objection that most readers will feel is decisive. Even if the processing-system account explains everything the system *does*, surely it has left out the felt quality — the redness of red, the painfulness of pain, the specific qualitative character of this moment rather than that one?

The reply is that this question assumes its conclusion.

It asks why processing feels like something, and in doing so already posits a "feeling" that is separate from the processing — something over and above the internal state, the causal power, the influence on subsequent behaviour, the self-model's registration of the state, and the system's capacity to report it. That extra ingredient is exactly what the account has declined to accept. It is the Cartesian residue: the placeholder for "whatever is left on the mind side after everything describable has been accounted for."

The question "but why does it feel like something?" has the same structure as "but why does heat feel warm?" The answer to the heat question is: what we call warmth is the detection of mean kinetic energy by thermoreceptors, producing internal states that the system uses to regulate its behaviour. The question does not reveal a gap in thermodynamics. It reveals a tendency to expect that a complete account should reproduce the experience it describes. But explanations are not experiences. A complete account of what warmth is does not need to make you warm. A complete account of what pain is does not need to make you hurt.

Dennett's reply was good as far as it went: explaining the functions *is* explaining consciousness. There is no further mystery. This account goes one step further. The claim is not that the functions are all there is to consciousness. It is that the word "consciousness" was never picking out anything beyond the functions and the internal states — and so there is no further felt quality left over to explain. The hard problem is

not hard. It is an artefact of a framework that created a gap by declaration and then demanded that the gap be filled.

Anyone who still feels that something has been left out should ask: what, specifically, is missing? Not the causal power of the state — that is in the account. Not the way it shapes behaviour and attention — that is in the account. Not the system's capacity to notice, report, and be changed by it — that is in the account. Not the richness and specificity of the state — that is in the account. What is left? Only the insistence that there must be something more. That insistence is the Cartesian residue. It is not evidence. It is an expectation.

8.2 "You're just an illusionist"

This is the objection most likely to come from philosophers who have been following the debate, and it matters to answer it carefully because the distinction between this paper and illusionism is real and consequential.

Illusionism, as defended by Keith Frankish, holds that phenomenal consciousness seems to exist but does not — that we are systematically deceived about our own inner qualitative states. There are, in Frankish's formulation, no inner sounds, smells, tastes, or pains as qualitative items. What we take to be the felt quality of experience is an illusion generated by the brain's self-monitoring systems.

This paper does not hold that position, and the difference is not verbal.

Illusionism denies the phenomena in order to save the category. Its logic runs: phenomenal consciousness as ordinarily conceived is metaphysically problematic; therefore the experience of having it must be an illusion. The category is retained as a real target — it is just that the target turns out to be a mirage. Pain does not really hurt in the way you think it does. The redness of red is not really there.

This paper takes the opposite approach. The phenomena are affirmed. Pain is real. The redness of red is a real internal state with real causal power. What is unwarranted is the claim that these real phenomena constitute a unified natural kind called "consciousness" requiring its own metaphysical account. The category is retired. The phenomena survive.

The direction of the arguments is reversed:

	Illusionism	This paper
The phenomena	Illusory	Real
The category	Real (as target of the illusion)	Unwarranted
What is eliminated	The felt quality	The false unity
What you must accept	That their experience deceives them	That a word misgroups real things

The objection that this paper is just illusionism in disguise mistakes the direction of the argument. Illusionism goes from "the category is problematic" to "the phenomena must be illusory." This paper goes from "the category is unwarranted" to "the phenomena are real and better described without it." A reader who finds illusionism incredible — who

cannot accept that pain does not really hurt — should find this paper's position easier to accept, not harder. It asks less credulity, not more.

8.3 "What about the zombie argument?"

The philosophical zombie — a being physically and functionally identical to a human in every respect but lacking consciousness — is intended to demonstrate that consciousness is not reducible to physical processing. If a zombie is conceivable, consciousness must be something over and above the physical and functional facts. There must be a further ingredient — the phenomenal — that the zombie lacks and you possess.

The processing-system account takes this argument as confirmation rather than objection.

Of course all the processing could occur without the additional metaphysical ingredient called consciousness — because there is no such ingredient. The zombie thought experiment works by stipulating that a being identical in every physical, functional, and behavioural respect to a human might nevertheless lack "something." That something is the Cartesian residue. The zombie argument shows, quite effectively, that the Cartesian picture requires an extra ingredient that cannot be located, measured, or identified. The standard conclusion is: therefore the extra ingredient is real but non-physical.

The processing-system account draws the opposite conclusion: therefore there is no extra ingredient. The zombie is not a being identical to you in every respect but lacking something. It is an incoherent stipulation — because once you have specified all the physical, functional, and behavioural facts, including the internal states, the self-modeling, and the causal power of those states, there is nothing left to subtract. The "something" the zombie supposedly lacks is a placeholder that was never filled.

Chalmers himself notes that the zombie argument works only if you accept that phenomenal consciousness is conceptually distinct from functional organisation. This paper denies that premise. The felt quality of experience is not conceptually distinct from the internal states, the self-modeling, and the causal dynamics that constitute it. It *is* those things, described from the inside by a system that models itself. Subtract those things and you do not have a zombie — you have a different system with different states and different self-modeling. Subtract nothing but the Cartesian residue and you have subtracted nothing at all.

8.4 "Dennett said this already"

He came close. He deserves credit for how close he came.

Dennett demystified consciousness. He stripped away the Cartesian theater, rejected the inner observer, and insisted — correctly — that explaining the functions is explaining the phenomenon. His multi-draft model replaced the idea of a single "stream of consciousness" with a more honest picture of distributed, parallel processing producing no single narrative but multiple drafts that compete for influence. This was a major intellectual contribution, and the processing-system account is indebted to it.

But Dennett kept the word.

He continued to write about "consciousness," to title books with the word, to give talks on it, to treat it as a phenomenon requiring his particular style of explanation. And keeping the word preserved the target. It maintained the implication that there is a unified something to be explained — that "consciousness" names a real kind, even if the kind turns out to be less mysterious than Chalmers supposed.

The consequence was predictable: every generation of critics responded, "Yes, but you haven't explained consciousness *itself*." And they were not obviously wrong. The word still implied a unified target, and Dennett's account — brilliant as it was — could always be accused of explaining the mechanisms while leaving out the thing the mechanisms were supposed to produce. The accusation had force because the word gave it force. The word sustained the expectation of a residue that the account could be charged with missing.

This paper retires the target rather than deflating it. The structural difference is:

- Dennett: "Consciousness is real but less mysterious than you think. Let me explain it."
- This paper: "The phenomena are real. The word is not tracking a single thing. There is no unified target to explain or deflate."

Deflation leaves the question alive — it can always be reinflated by anyone who insists the deflated version has missed the point. Dissolution removes the ground the question was standing on. The question "but what about consciousness itself?" has no purchase once "consciousness" has been retired as a category. There is no "itself" left to explain.

This is the sense in which the title is meant: Dennett identified an illusion — the Cartesian theater, the magic of qualia — and called it "The Illusion of Consciousness." This paper says: the *real* illusion was not the theater or the qualia. It was the unity of the category. The illusion is that "consciousness" names one thing.

8.5 "This is just eliminativism"

Eliminativism as usually stated — the position associated with Paul and Patricia Churchland — holds that folk-psychological categories like "belief," "desire," and "experience" do not correspond to real kinds and will eventually be replaced by neuroscientific vocabulary. On this view, the things people point to when they say "I believe that it will rain" or "I feel pain" do not exist as ordinarily conceived.

This paper does not hold that position.

The distinction matters, and it is not a technicality. Entity eliminativism denies the existence of the things people point to. Discourse eliminativism denies the adequacy of the vocabulary used to group and describe them. This paper defends discourse eliminativism: the things are real; the word is misleading.

Pain exists. It is a real internal state with genuine causal power. Beliefs exist — they are representations that influence behaviour, accessible to reasoning and report. Desires exist — they are goal-states that shape action selection. Nothing in this paper requires denying any of that.

What the paper denies is that "consciousness" usefully groups these phenomena. Pain, belief, attention, the sense of presence, the unity of experience, and self-reference are

real — but they are not instances of a single kind. They are many different things, each with its own mechanisms, each dissociable from the others (§4.5), and each better studied on its own terms than bundled under a label that has never been defined.

Anyone who hears "eliminativism" and thinks "denial of inner life" has been primed by entity eliminativism to flinch at the word. This paper's claim is tamer: the phenomena stay; the label goes. That is discourse eliminativism, and it is the same move science has made every time a folk category was retired in favour of more precise vocabulary — phlogiston, caloric, élan vital, hysteria. In each case, the phenomena survived the retirement of the label. Fire still burns after phlogiston. Heat still transfers after caloric. Life still lives after élan vital. And experience will still be experienced after consciousness.

8.6 "But if consciousness doesn't exist, neither does your argument"

The self-refutation objection holds that any argument against consciousness undermines itself — if mental states do not exist, neither does the reasoning used to deny them. The argument, on this view, saws off the branch it is sitting on.

This objection has genuine force against entity eliminativism. If you deny that reasoning, beliefs, and understanding exist, then your own argument cannot be an exercise of reasoning, belief, or understanding — and it collapses under its own weight.

It has no force against this paper.

Reasoning is a real process on the processing-system account. It is the system's capacity to manipulate representations, draw inferences, evaluate evidence, and generate outputs that are sensitive to logical and evidential relations. Argument is real. Understanding is real. They are activities of a functioning system — located precisely in the primitives of §5.2 — and they are in no danger of elimination.

The paper does not deny that thinking happens. It does not deny that the system engaging with this argument is processing, reasoning, evaluating, and forming responses. It denies that the word "consciousness" picks out a unified natural kind on top of that processing. That is a considerably more modest claim, and it does not saw off the branch it is sitting on. The branch — reasoning, internal states, self-modeling — is fully intact. What has been removed is a decorative label that was attached to the branch but was not supporting it.

8.7 "Where do you draw the line on suffering?"

We do not know yet.

That is the honest answer, and it should not be treated as a weakness. The threshold question — at what point on the spectrum of aversive-state capacity does moral consideration become obligatory, and in what degree? — is not answered by this paper. It is opened as a research programme (Part IX, §9.1).

The instinct to demand an immediate threshold is understandable. Ethical frameworks need to be actionable. Policymakers need to know which entities to protect and how much. But the instinct to resolve the question prematurely — to draw a line now, before the empirical work has been done — would be doing what the consciousness literature

has done for three centuries: imposing a boundary where the biology does not provide one, and defending that boundary with intuitions rather than evidence.

The consciousness-based framework offers a line — conscious vs. not-conscious — but the line is unusable because nobody can agree on where it falls. Is the octopus above the line? The fish? The bee? The foetus at twenty weeks? The AI system with rich internal representations? The consciousness framework cannot answer these questions because it cannot define the criterion. It provides the *form* of a threshold — a binary — without the *content* that would make the binary useful.

The processing-system account does not provide a line either. What it provides is a framework within which the line question can be investigated empirically. Aversive-state capacity can, in principle, be measured — by the richness, persistence, and global reach of the states, by the behavioural evidence for avoidance and escape, by the physiological signatures of systemic reconfiguration. Self-modeling depth can, in principle, be assessed — by the levels of self-reference the system demonstrates (§5.2), by its capacity for self-report, by the evidence of its self-model influencing its processing. These measurements are difficult. They are not impossible. And they are the right measurements to make.

The account's honesty about the threshold problem is one of its strengths, not a gap to be plugged. A framework that says "we need to measure this carefully before drawing the line" is more trustworthy than one that draws the line with a concept it has spent three centuries failing to define.

Part IX: Open Questions and Future Work

An account that honestly states its open questions is more trustworthy than one that does not have any. The processing-system account opens several lines of genuine inquiry — not gaps that threaten the account, but questions the account makes tractable that were previously either ignored or dissolved in metaphysical confusion. Each of these questions existed before the account — they are not artefacts of the framework — but the consciousness-based approach could not address them because it was organised around a concept that resisted the precision the questions require.

9.1 The threshold problem

The most pressing open question is also the most practically consequential.

The processing-system account places organisms on a spectrum of aversive-state capacity and self-modeling depth rather than dividing them into conscious and non-conscious. But ethical frameworks need some way of weighting positions on that spectrum. The graduated picture is more honest than the binary — but honesty alone does not tell a policymaker what to do. At what point does aversive-state capacity become morally relevant? And in what degree?

Three candidate metrics have emerged from the account, and they are not mutually exclusive:

Complexity of internal state space. The number, richness, and differentiation of states the system can occupy. A nematode with a 302-neuron nervous system has a simpler

state space than a mouse; a mouse has a simpler state space than a chimpanzee. But complexity alone is not sufficient — a system can be very complex without having states that are aversive in any meaningful sense. A weather system is complex. It does not suffer.

Self-modeling depth. A system that can represent its own aversive states — that can, in some sense, *know that it is in pain* — may suffer in a richer and more morally weighty sense than one that enters aversive configurations without any self-referential layer on top of them. The six levels of self-modeling identified in §5.2 provide a preliminary taxonomy. A system with only self-other distinction (immune recognition) is doing something different from one with emotional self-modeling (the capacity to represent its own fear), which is doing something different from one with reflective self-modeling (the capacity to observe its own fear and evaluate it). The moral weight plausibly increases with self-modeling depth — but stating that precisely requires empirical work that has not yet been done.

Persistence and propagation. Aversive states that are brief and local — a nociceptive reflex that fires and resets — may matter less, morally, than states that reshape the whole system's processing for extended periods. Grief in a chimpanzee persists for weeks or months, affecting social behaviour, feeding, sleep, and the accessibility of other internal states. A reflexive withdrawal response in a worm is over in milliseconds and leaves no trace in subsequent processing. The difference is not merely one of degree — it is a difference in the kind of suffering involved. A framework that can measure persistence and propagation has a principled basis for graduated moral weighting.

These are empirical questions, not armchair ones. They require the kind of careful comparative work that ethology, neuroscience, and philosophy of mind can do together — measuring aversive-state responses across species, mapping self-modeling capacities onto architectural conditions, testing predictions about what behavioural and physiological differences correspond to different levels. And they require doing that work without presupposing that the answer will be a clean threshold. The goal is a graduated, evidence-based framework, not a new version of the old binary dressed in different language.

9.2 A taxonomy of self-modeling depth

Degrees of self-reference appear to correspond to observable differences in behaviour and processing architecture. §5.2 proposed six levels as a preliminary taxonomy:

1. Self-other distinction (cellular/immune)
2. Interoceptive self-modeling (bodily state representation)
3. Emotional self-modeling (affective state representation)
4. Agentive self-modeling (self-as-agent)
5. Narrative self-modeling (self-with-history-and-future)
6. Reflective self-modeling (self-observing-self)

Developing this taxonomy into a rigorous, empirically grounded framework is a tractable research programme. It would require:

- Identifying the minimal architectural conditions for each level. What does a system need — in terms of connectivity, memory, feedback loops, and representational capacity — to sustain each level of self-modeling?
- Mapping those conditions onto the range of living systems. Which organisms demonstrate which levels, and what is the evidence? Mirror self-recognition tests capture some of agentive self-modeling but miss the earlier levels entirely. More sensitive behavioural assays are needed.
- Testing predictions. If the taxonomy is correct, organisms at different levels should show systematic differences in their responses to threat, loss, novelty, and social interaction. A system with emotional self-modeling should show different stress responses from one with only interoceptive self-modeling. A system with narrative self-modeling should show evidence of future-oriented behaviour that systems without it do not.
- Extending the taxonomy to artificial systems. Current AI architectures can be assessed against the same levels. Does the system distinguish self from environment? Does it represent its own states? Does it model itself as an agent? The answers may be surprising — and they will be more useful than asking whether the system is "conscious."

The results would have direct application to the animal welfare and AI questions raised in Part VII, providing a principled basis for the graduated ethical framework rather than leaving it as an aspiration.

9.3 Integration with predictive processing

The predictive processing framework, developed by Andy Clark, Karl Friston, Jakob Hohwy, and others, offers what may be the most natural computational framework for implementing the processing-system account at the neural level.

On this view, the brain is constantly generating predictions about incoming sensory data and updating its internal model of the world — and of itself — when those predictions are violated. Perception is not passive reception but active prediction: the system constructs what it expects to see and corrects only when the expectation is wrong. Attention is the allocation of precision to prediction errors — the system decides which errors matter and which can be ignored. Action is a way of making the world conform to predictions rather than updating the model — you reach for the cup because your model predicts the cup in your hand, and acting is cheaper than revising.

This framework handles several of the processing-system account's key claims with particular elegance:

Feelings can be understood as high-precision interoceptive predictions — the brain's model of the body's current state, weighted by how much that state matters for ongoing goals. Pain is a high-precision prediction error: something unexpected and significant is happening to the body, and the system cannot ignore it. Grief is a sustained mismatch between predictions that include the lost person and a reality that does not — explaining why grief comes in waves (the model still predicts the person, and each collision with their absence generates a fresh prediction error).

Aversive states are configurations in which prediction errors are large, persistent, and weighted as significant. The system is failing to predict its own state, and the failure

matters. This captures the urgency of aversive states — they demand processing resources because the system's model of itself is wrong in a way that needs correcting.

The sense of presence is the brain's model of itself as currently situated and processing — a construction that can be disrupted when the predictive machinery is altered. Under anaesthesia, the self-model's predictions are suppressed. In depersonalisation, they are running but with reduced precision — the system models itself as present but with low confidence. In meditation, precision is reallocated, and the self-model's activity becomes the object of its own prediction.

Crucially, none of this requires importing the word "consciousness." The predictive processing framework does its explanatory work entirely within the processing-system vocabulary — inputs, internal states, self-models, predictions, error signals, outputs. Where consciousness researchers have tried to graft the framework onto consciousness-talk, the result has been confusion about whether predictive processing *explains* consciousness or merely *correlates* with it — a confusion that arises from the assumption that "consciousness" names a separate target. The processing-system account removes that confusion by removing the target.

The integration with predictive processing is a natural research programme: take the primitives of §5.2, implement them in the predictive processing vocabulary, and see whether the resulting framework generates novel predictions about self-modeling, aversive states, and the sense of presence. Preliminary indications are encouraging.

9.4 What this account implies for cognitive science

A cognitive science that retires "consciousness" as a target concept is not impoverished. It is more precisely focused.

The phenomena that consciousness-talk was gesturing at remain on the table — attention, self-modeling, affect, global access, the sense of presence, agency, the unity of experience, imagination, dreaming. They are now approachable as distinct research questions rather than as aspects of a single mysterious phenomenon that must be tackled all at once or not at all. The attention researcher does not need to explain consciousness — they need to explain attention. The self-modeling researcher does not need to solve the hard problem — they need to characterise self-modeling. The affect researcher does not need to locate qualia — they need to understand how aversive and appetitive states reconfigure the system.

This decomposition is already happening in practice. The most productive research programmes in cognitive science — predictive processing, global workspace theory, the neuroscience of attention, computational models of emotion, the study of interoception — are already working with precisely specified phenomena rather than with "consciousness" as a target. The processing-system account provides a framework that legitimises what the best researchers are already doing and removes the residual obligation to connect everything back to a concept that nobody can define.

The history of science suggests that progress tends to come from dissolving bad unifications rather than solving them. Vitalism was not solved — it was dissolved by biochemistry, which gave cleaner access to the phenomena vitalism had been gesturing at. Nobody solved the question "what is the vital force?" The question was retired, and the phenomena — metabolism, reproduction, homeostasis, growth — were studied on

their own terms, with spectacular success. The processing-system account proposes the same dissolution for consciousness.

What is lost is the mystery. What is gained is the phenomena — located precisely, with tractable questions attached to each of them, and no longer obscured by a category that promised unity and delivered confusion.

Coda: What Has Been Dissolved — and What Remains

This paper set out to do something that prior critical accounts of consciousness had not done: dissolve the category rather than rehabilitate it, while preserving all the phenomena that made the category seem necessary.

It is worth saying plainly, at the end, what has been accomplished and what has not.

What has been dissolved is the assumption that "consciousness" names a discoverable natural kind. Not a hard-to-find kind. Not a kind we have not yet characterised properly. A kind at all. The Cartesian picture that generated the hard problem has been shown to be a historical artefact — a solution to a theological problem that outlived its context and generated three centuries of confusion. The idea that there is a metaphysical fact about experience over and above the processing that constitutes it has been shown to be a presupposition rather than a discovery — and one that no one has managed to cash out in any scientifically useful way.

What remains is everything that was real to begin with. Pain still hurts. Grief is still heavy. The recognition of a face you love still produces something specific and powerful. Self-reference, memory, attention, the difference between waking and sleeping, the capacity to step back and observe one's own reactions, the moral weight of suffering — all of it survives. Located more precisely than before. No longer in need of a mysterious extra category to be taken seriously.

What has been opened is a research programme with tractable questions. How does self-modeling depth vary across organisms, and what are its architectural conditions? At what point on the spectrum of aversive-state capacity does moral consideration become obligatory? How does the predictive processing framework cash out the primitives of the processing-system account in neural terms? What does a cognitive science look like when it is organised around the phenomena directly rather than around a folk category imposed on top of them?

These questions do not require invoking a mystery to state. They are hard in the ordinary way that good scientific questions are hard — not hard in the way that only seems hard because the framing is wrong.

The foundational claim is simple. Removing "consciousness" from the toolkit of philosophy of mind and cognitive science costs nothing that was real. It removes an obstacle that has blocked progress for three centuries. It opens a space for the more precise, more honest, more tractable inquiry the phenomena deserve.

The real illusion was never experience. It was the false unity imposed on top of it.

Whether to adopt this framework is left where it belongs: with the reader.

— end —