

Selection and Instantiation

Where physics needs the indexical, why that is not the same as needing consciousness, and what survives of the idea that a theory can eat its own evidence.

ABSTRACT

Physics describes the world exclusively in terms of structure and dynamics, and Newman's 1928 result shows that purely structural knowledge reduces to a cardinality claim plus mathematics. Set beside the inner case — intrinsic character given, structural role not — this yields an exact inversion of epistemic profiles and locates the explanatory gap in the form of physical theory rather than in the deliverances of introspection. Working from that carving, this paper defends eight results. Subjectivity is not indexicality: self-location is selection from an antecedently given domain, and indexicality without phenomenality is realised in commercial navigation systems. Cognitive instability formalises as a self-applying assertability audit — an undercutting defeater with a self-referential trigger — which binds epiphenomenalism, stalemates honestly with illusionism, and survives amputation of the anthropic machinery it appeared to need. Constitution delivers reference without information. Newman applied to integrated information theory yields a dilemma between notational and metaphysical intrinsicity. Combination is placed outside physics by an argument from physics. Felt passage decomposes into five components, of which only flow survives. And the undetectability results in three separate literatures are shown to be one result, which inverted becomes a science-eligibility test whose consequence is that eligible theories are mutually indiscriminable. The hard problem is not solved here; §XV states precisely what is not.

Merlin · 24–25 August 2026 Eight results · three instruments · §I now conditional on OSR

I · THE CARVING

Structure without intrinsic, intrinsic without structure

The vague form of the inner/outer problem and the sharp form are different animals. This is the sharp form.

Physics describes the world exclusively in terms of structure and dynamics: a state space, a symmetry group, an evolution law. Relations among relata, with the relata specified only by their position in the relations. Russell saw this in *The Analysis of Matter* (1927) — physics hands over the causal skeleton and is silent on the intrinsic nature of whatever does the causing. Eddington's version: a schedule of pointer readings.

It has formal teeth. Newman's 1928 objection to Russell: if knowledge of the external world is purely structural, then for any set of the right cardinality one can *define* relations on it satisfying that structure. "The world has structure S" degenerates into "the world has at least this many things," plus mathematics. Pure structuralism about the outer is very nearly vacuous — a theorem, not a mood, and the hardest constraint in the area.

Set the inner beside it and the asymmetry is exact:

OUTER structure given, intrinsic nature inaccessible
INNER intrinsic nature given, structural role inaccessible

REVISED · §VII

As first stated this treats "inner" as a *domain*. It isn't. §VII establishes that the zone of guaranteed access is a razor edge — a single occurrent demonstrative carrying no classificatory content — and that everything else about one's own mind is ordinary fallible description. Read the profile above as true of that edge only. The rest of inner reality is known roughly the way the weather is.

The one thing in the universe whose intrinsic character is directly presented is the one thing physics cannot locate; everything physics does locate is known only up to isomorphism. This is why Russellian monism keeps being independently rediscovered by physicists: physics leaves a hole where intrinsic natures should be, consciousness is a fact with no address, and perhaps there is one hole. The bill arrives as the combination problem, posed by James in 1890 and still unpaid — micro-experiential facts do not obviously sum to a unified macro-experience.

THE EXPOSURE: ONTIC STRUCTURAL REALISM

The step from "pure structure is vacuous" to "something non-structural obtains" is where this paper is most vulnerable, and the position that denies it should be named rather than passed over. Newman's result is standardly taken to be decisive against *epistemic* structural realism — Demopoulos and Friedman showed that on the Ramsey-sentence version "only cardinality questions are open to discovery," every other claim following a priori once the observational consequences are fixed. Ladyman introduced the *ontic* version partly in response, and Ladyman and Ross state the escape directly: "Since ontic structural realism is not formulated in these terms, the Newman problem does not arise for ontic structural realism. In particular, we will eschew an extensional understanding of relations without which the problem cannot be formulated." They add the physics — entanglement gives relations that "do not supervene on the intrinsic properties of their relata" — and conclude that "things in themselves and qualia are idle wheels in metaphysics."

So §I's claim is true of the extensional reading of relations and question-begging against the intensional one. Stated without OSR in view, it asserts precisely what OSR denies.

The repair is available, and it is §IX's fork one level up. OSR escapes by taking relations intensionally — modally robust rather than sets of tuples. Ask what grounds the modal robustness and there are two answers. Either it is primitive modal or causal force, in which case a non-structural primitive has been posited and the disagreement with Russellian monism is about where the primitive sits rather than whether there is one. Or nothing grounds it, and the extensional reading returns with Newman behind it. That is notational versus metaphysical intrinsicity in different clothing, and §IX's second instrument — *declare your primitive* — runs on OSR unmodified.

This does not refute OSR and is not offered as doing so. Esfeld's moderate metaphysics of relations accepts relata without intrinsic natures and is a live position. What it does is convert §I from an assertion into a conditional with a stated price. Worth noting that the price is one an OSR proponent has already acknowledged: Esfeld grants that the structures "are usually conceived as including a primitive modality" and that "it has not been spelled out as yet what exactly that modality amounts to."

And the reason none of this is yet a science is narrower than usually admitted. It is not that we lack a theory. **We lack an observable.** No experiment distinguishes a conscious system from a functionally identical unconscious one. Nearly every downstream difficulty descends from that single deficit.

The render is real, and proves less than it seems

The sun you see is eight minutes old. The moon is roughly the colour of worn asphalt and is rendered bright because it hangs against black. Sharp colour vision spans about two degrees — a thumbnail at arm's length. There is a hole where the optic nerve exits and you never see it. The eyes saccade three or four times a second with vision suppressed throughout, so you are blind for something like an hour a day and have never noticed a gap. The seamless field in front of you exists nowhere as data. The retina itself is not an instrument attached to the brain but an outgrowth of it, the only central nervous tissue visible from outside a person.

Kant's limit is the one that holds: you cannot get behind the render to check it, because checking is itself a render. The thing-in-itself is not a hidden object behind the moon; it is the name for where the question runs out.

TRAP

The standard next step — *therefore you never see the world, only your brain's representation* — uses the retina, optic nerve and cortex, objects known entirely through perception, to conclude that perception does not reach objects. Every fact about the optic nerve was obtained by someone looking at one. The argument eats its own evidence. This is the first appearance of the structure that becomes the instrument in §VI.

What the physiology does establish is Helmholtz's point (1860s): the retinal image is fatally ambiguous, since infinitely many scenes project the same pattern, so vision solves an underdetermined inverse problem by inference. Seeing is a hypothesis continuously tested, which is exactly why illusions work — they are well-formed bets that happen to be wrong, and only a system guessing intelligently can be fooled at all. Charles Bonnet syndrome suggests the retina is the *occasion* for seeing rather than its seat: cut the feed and the machinery keeps rendering.

The residue is Leibniz's, from the *Monadology*: enlarge the brain to the size of a mill and walk in, and you find parts pushing on parts. Nowhere in the machinery is the perception. Total dependence and missing explanation, simultaneously.

Where the first person is actually load-bearing

GROUND CLEARING

The measurement problem is not the consciousness problem. Von Neumann (1932) showed the cut between quantum system and classical apparatus can be placed anywhere along the chain — particle, detector, retina, cortex — with identical predictions. The mobility of the cut is a theorem, and what it establishes is that physics gives no reason whatsoever to place the cut at consciousness. London and Bauer (1939) and Wigner (1961) placed it there anyway; Wigner later abandoned the position.

Decoherence does the work instead: environmental entanglement destroys interference between pointer states on timescales around 10^{-31} s for a dust grain in air, and einselection picks the preferred basis out of the interaction Hamiltonian rather than out of anyone's attention. But the limit is real — decoherence converts a coherent superposition into an improper mixture. It explains why you never see interference; it does not explain why there is one outcome. In Bell's phrasing, it gives you *and* where you wanted *or*.

FOUR GENUINE CONTACT POINTS

Everett and self-location. The dynamics is unitary and deterministic, so the only uncertainty left anywhere is indexical: after branching and before observation you do not know which branch you are in. Not a fact about the wavefunction, which contains all of them — a fact about which copy you are. Deutsch (1999) and Wallace derive Born decision-theoretically from this; Sebens and Carroll do it explicitly through self-locating uncertainty. Whether the derivations succeed is contested. The architecture is the point: in the most austere realist reading of quantum mechanics, the empirical content of the theory is grounded in a first-person self-locating fact.

Wigner's friend, modern form. Frauchiger and Renner (2018): agents modelled as quantum systems, reasoning about other agents' quantum reasoning, derive contradictory conclusions about a single run. One of universal applicability, cross-agent consistency, or single-outcome definiteness must go. Bong et al. tested a local-friendliness version photonically in 2020. Observers here are formal objects, and the arithmetic breaks when perspectives are composed.

QBism. Fuchs, Mermin and Schack take the quantum state as an agent's personal degree of belief about their own future experience and the Born rule as a normative coherence condition. Wigner's friend, EPR nonlocality and the measurement problem all evaporate — at the price that it becomes very hard to say what the theory is about. The only serious programme that puts the experiencing agent into the formalism deliberately.

The missing now. Nothing in physics corresponds to the present moment. Simultaneity is frame-relative; the block has no privileged slice; Weyl's formulation is that the objective world simply *is*, it does not *happen*. Yet the most vivid feature of inner reality is that it is now, and that the now moves. Worse for the arrow: the fundamental laws are time-symmetric to excellent approximation, so the directionality of inner life descends from no law but from the Past Hypothesis, a low-entropy boundary condition at the beginning. **Your inner arrow of time is a cosmological initial condition, felt from inside.** And it runs both ways: cosmologists reject models on Boltzmann-brain grounds, which is the first-person datum *my memories are veridical* being used as evidence against physical theories.

Physics never once needs consciousness to make anything happen. It repeatedly needs the indexical — which branch, which moment, which observer, which copy — to make its own content well-defined.

IV · REGISTER

The register, and where each question stands

01 Is subjectivity just indexicality wearing a costume?

WORKED · §V

Resolved negative. Selection versus instantiation. The knowledge argument takes collateral damage.

02 Does Newman's theorem kill computational theories of consciousness?

WORKED · §IX

Run. Result: a dilemma — notational intrinsicality lets Newman consume IIT's predictive content; metaphysical intrinsicality blocks Newman but axiomatises the gap. Generalises to any theory answering §I with a formal quantity.

03 Can cognitive instability be made a formal constraint?

WORKED · §VI

Partially. Yields an assertability audit with known limits. The statistical limb was amputated.

04 Is felt passage identical to the entropy gradient, or parasitic on it?

WORKED · §XII

Run. Parasitic, component-specific: direction reduces to the gradient (inversion argument), memory to records, presentness to indexicality, duration to neural timescales. Only flow survives, and its reality is what the remaining dispute is about. Inverted phenomenal time is *not* truth-apt as a comparative claim — which is what does the reducing.

05 What would an observable look like — and has AI removed the last crutch?

WORKED · §XIII

Run. The three undetectabilities are one, and the one is §VI's coupling criterion arriving from three literatures. Inverted, it becomes a science-eligibility test — the first constructive instrument here. AI's role stands: it strips the similarity heuristic rather than creating a new problem.

06 Could the combination problem be attacked with renormalization-group tools?

WORKED · §X

Run. It was a category error, and establishing that *is* the result: Russellian monism gives the phenomenal no coupling, so RG cannot see it, and the inapplicability turns out to be forced rather than accidental. Physics contributes one negative finding — universality strips the genetic argument of independent force.

07 Does constitution generalise, or is "knowing" two relations?

WORKED · §VII

Arose as the epiphenomenalist's escape from §VI. Resolved: constitution is a genuine relation that delivers reference without information, and does not reach the case it was recruited for. Consequence in §VIII.

V · RESULT ONE

Subjectivity is not indexicality

Negative verdict, with a consolation prize for the deflationist that turns out to be worth more than the identification would have been.

THE DEFLATIONARY CASE, AT FULL STRENGTH

Lewis's two gods inhabit a world. One lives on the tallest mountain and throws thunderbolts, the other on the coldest and throws manna. Each knows every proposition true at that world — the description is complete. Neither knows which god he is.

That is Mary's room in structure: complete propositional knowledge, something still missing. And nobody concludes from the two gods that non-physical facts exist. Lewis makes self-location the self-ascription of a property rather than belief in a proposition — a set of centred worlds, not a set of worlds — and separately analyses Mary's gain as ability rather than information. The moves rhyme because they are the same move. So the deflationist has a real point: indexicality is an uncontroversial, fully naturalistic case where the complete third-person description leaves something out, with no ontological consequence, and that is exactly the epistemic signature the knowledge argument treats as diagnostic of dualism.

WHY THE IDENTIFICATION NONETHELESS FAILS

INDEXICAL FACTS are selection facts. They pick one item from a domain already given. "Which branch" presupposes the branches.

PHENOMENAL FACTS are not. The question is not which experience among the available ones is mine — that would be indexical, and answerable. It is why any position is occupied from the inside at all.

The counterexample is cheap and industrial. A robot doing SLAM — simultaneous localisation and mapping — maintains a world model with a marked position in it, updates the marker under uncertainty, and acts on the indexical result. That is genuine self-location, on hardware you can buy, with nobody home. **Indexicality without phenomenality is not a thought experiment; it is a product category.**

THEN WHY DOES PHYSICS KEEP NEEDING IT?

Because physics is produced by subsystems measuring from inside the system they describe. Everett needs "which branch" because the observer is within the wavefunction. Relativity needs frames because the measurer is on a worldline.

Anthropics needs "which observer" because the theorist is a sample from the population the theory quantifies over. The recurrence is evidence of **embeddedness, not interiority**. Ismael's programme is exactly this: the self-model as a physical system's internal representation of itself, and how that generates the appearance of a privileged standpoint.

THE RESIDUE WORTH KEEPING

The identification fails, but the knowledge argument takes real damage on the way out. We now hold an undisputed case — the two gods — proving that *there are facts unavailable from the complete third-person description* does not entail *there are non-physical facts*. The anti-physicalist must now explain why Mary is a different kind of case. That burden is new, and selection versus instantiation is where to make them stand.

WHY SLAM IS NOT QUESTION-BEGGING

The obvious reply is to deny the case: insist the robot has some minimal experience. Cheap to say and hard to refute — so the burden must be discharged without appeal to whether the robot is conscious. It can be, because the reply's own

commitments do the work. There are three ways for the robot to have experience and all three leave the result standing. Note that the second horn covers every theory identifying experience with a structural or functional feature, and does so without needing §IX — so its reach does not depend on the instrument sweep §XVI records as outstanding.

SUBSTRATE-LEVEL (the Russellian line) – the robot's experience is in its matter's intrinsic natures, so a rock has it too. Then its experience is *independent* of its self-location, and the self-location is still fully accounted for with no reference to it. The case goes through.

ORGANISATION-LEVEL (the structural line – IIT, global workspace, higher-order, attention schema) – experience is identified with a structural or functional feature. Then §IX's dilemma applies to the Φ -based version; and more simply, and for all of them, a structural fact is third-personally available, so the robot's experience cannot be the further thing that makes its self-location phenomenal rather than merely indexical. It is the self-location, redescribed.

FROM SELF-LOCATION AS SUCH – indexicality suffices for phenomenality. The nearest live position is self-representationalism, on which a specific same-order self-representational structure suffices; but that is not bare indexicality, and SLAM's marker is not a self-representation in the required sense. Anyone who does take bare self-location to suffice makes phenomenality cheap rather than indexicality expensive: the identification collapses the wrong way.

Load-bearing note. §VII offers an account of what happens at the moment of givenness as a mechanism for this result, and flags a live tension between that mechanism and §VIII. The architectural repair given there does the same non-question-begging work as the trilemma above and should be treated as load-bearing; the bare appeal to SLAM without either is the weakest branch, not the safe one.

VI • RESULT TWO

The assertability audit

What survives of the intuition that a theory destroying its own evidence is out. Less than promised, but real, formal, and with known limits.

THE CORRECT LOGICAL TYPE

Not self-contradiction — the theories are consistent. Self-undermining in the evidential sense, and the technical home is Pollock: an **undercutting defeater**, which attacks the link between evidence and conclusion rather than arguing against the conclusion. Rebutting defeaters give reason to think T false. Undercutters give reason to think your reason was never a reason. The term "cognitively unstable" is Carroll's (2020), which cites related discussion in Albert (2000, p. 116); the formal treatment below is Elga's (2025), and the definition that follows is corrected against it.

Let T be a theory, E your total evidence, R the proposition that E was produced by a process tracking the relevant facts.

T is cognitively unstable when $P(R \mid T) \approx 0$, your only route to T ran through E , and T leaves E probabilistically inert – $P(E \mid T) = P(E \mid \neg T)$, so no likelihood ratio survives the conditioning.

Conditioning on T then dissolves the support for T – a fixed-point failure.

FALSIFIED IN REVIEW • CLAUSE THREE

As first stated the third clause read *and E's evidential force depended on R*. Elga (2025) refutes it with a worked case. In Labelsramble you read a pill-bottle label reporting that the drug you have taken causes hallucinations that replace pill-bottle labels. $P(R \mid L) \approx 0$, your only route to L runs through that impression, and yet $P(L \mid E) = 1$ — the impression is decisive evidence for the very hypothesis that impugns it. His general statement is correct and the clause as written is not: "the deliverances of a faculty can simultaneously be evidence that the faculty is working improperly and also that it has happened to be correct in the present instance." Unreliability is not severance. An untrustworthy process can still produce an extreme likelihood ratio — and in his third version the ratio inverts, so the same fishiness becomes "almost decisive evidence that you did *not* take Labelsramble." The machinery is ratio-driven throughout.

So the audit does not run on distrust. It runs on **inertness** — which is the narrower claim, and the one the verdicts below were always tracking.

WHAT IT BINDS

TARGET	VERDICT	REASON
Epiphenomenalism	Lethal	The report "I am conscious" is caused wholly by physical states and would be produced verbatim in a world with no experience. Belief severed from truthmaker: causal inertness makes report, belief and utterance identical across worlds with and without experience, so $P(E \mid T) = P(E \mid \neg T)$ exactly — posterior odds equal prior odds, for every prior. <i>Not</i> the Boltzmann structure: there, Elga's case is that the evidence "was antecedently much more likely to arise if LargeBB is true than if not," so a ratio survives and the channel is noisy rather than cut. Epiphenomenalism is the degenerate point of Elga's own model, not an instance of the case he dissolves.
Simulation hypothesis	No purchase	Simulation is metaphysics, not scepticism. Beliefs about chairs are true and chairs are made of bits. Only the deceptive variants engage, at which point it has become the Boltzmann argument again.
Illusionism	Stalemate	Introspective reports are caused by representational states that misrepresent, so the causal chain to the utterance is intact. The undercutter needs a broken chain; there isn't one.

WHAT THE INSTRUMENT OWES ELGA

His target is the argument that instability is *inevitable*, and against that he is right; he says as much — "we needn't do so in order to get evidence that the above argument (that instability is inevitable) is no good." The audit never claimed inevitability. What his model supplies is the missing boundary condition: **instability is illusory wherever a likelihood ratio survives conditioning, and real where the theory pins that ratio at one**. Run his machinery over the three targets and it reproduces the table above — lethal for epiphenomenalism, no purchase on simulation, stalemate with illusionism. The instrument and the dissolution turn out to be the same instrument.

Which also explains why illusionism only stalemates rather than being dispatched. Introspective reports are caused by representational states that misrepresent, so theoretical reasoning about physical facts — the truthmakers — stays at least weakly truth-sensitive. The ratio is not one. Illusionism sits inside the region Elga dissolves, which is precisely why the audit cannot bind it, and why the stalemate is honest rather than a hedge.

The epiphenomenalist's remaining move is to deny the premise: the belief is tied to qualia by constitution rather than causation, which would restore a ratio. That is the hatch, and §VII closes it — constitution secures reference with no content, and the channel breaks at the first inferential step, which is exactly where the ratio would have had to come from.

THE REFERENCE CLASS PROBLEM IS STRUCTURAL, NOT TECHNICAL

Ordinary Bayesian inference has a physical process you can point at, specified independently of the hypotheses. Anthropic reasoning has none. **Nothing draws you**. No stochastic mechanism assigned you to being this observer, and no law supplies a measure over observers. The sampling is a fiction introduced to license the inference, and the reference class is that fiction's free parameter. The question is not whether the number can be pinned down; it is whether there is a number.

Cosmology has already run the experiment. The measure problem in eternal inflation is the reference class problem in a lab coat: regulating an infinite set requires a cutoff, there is no canonical one, and the *Boltzmann brain verdict itself flips* with the choice. Twenty years, by people who wanted it to exist.

The steelman is real and should be recorded. SIA cancels the reference class dependence in the combined update, killing the Doomsday shift exactly. Neal's full non-indexical conditioning eliminates the class by conditioning on total evidence read non-indexically — and fails informatively, since in a large universe someone with your exact evidence exists under

every hypothesis, so the likelihoods go to one and you learn nothing. Best of all is Armstrong's anthropic decision theory: SSA and SIA correspond not to rival likelihood facts but to rival value systems, average versus total, and once you fix what you care about the decision is determinate with no class ever specified. But note what that concedes — determinate decisions purchased by abandoning determinate credences. A dissolution, not a solution. Fatal here, because the audit was supposed to constrain belief.

And it meets Result One: a reference class is a specification of the domain. Self-location is selection *from* a domain and cannot generate one. Anthropic reasoning infers the structure of the domain from the position — the inference run backwards.

CORRECTION

I predicted the two questions would collapse into a single obstacle. Wrong. They *partially separate*, and the separation is the result: cognitive instability has a statistical limb, which dies with the reference class, and a causal limb, which never needed anthropics at all.

AMPUTATE THE ANTHROPIC LIMB

The statistical limb asks whether you are a typical observer. It needs a measure that does not exist. The causal limb asks whether this belief is counterfactually connected to its own truthmaker, or would have arrived identically in the absence of what it is about — a claim about the causal structure of one system. No population, no sampling, no class. And the epiphenomenalism argument only ever used the second. The Boltzmann brain case, which looked like the flagship, is the member of the family that requires the broken machinery. The philosophers were on firmer ground than the cosmologists the whole time.

IT IS NOT RELIABILISM

Reliabilism is a first-order thesis about beliefs applied by an assessor standing outside: does this process track truth? It requires an external vantage. The audit requires none, because it never compares the belief to the world — it feeds the theory its own content.

T entails that beliefs of type B are causally insensitive to their truthmakers.
"T is true" is a belief of type B.
∴ T entails the insensitivity of the only belief that could carry it.

Nothing there asks whether T is *actually* true or the process *actually* reliable. The steelman — that this is reliabilism evaluated under a hypothesis — is fair, but the work is done by the self-reference, which reliabilism does not have.

THE TEST THAT WOULD EMBARRASS IT

This schema has a disreputable history. Haldane, 1927: if my mental processes are wholly determined by the motions of atoms in my brain, I have no reason to suppose my beliefs true, and hence no reason to suppose my brain composed of atoms. Plantinga industrialised it against naturalism. If the audit is sound, does it validate Plantinga?

It does not, and why not is the most useful thing here. Naturalism plus evolution *supplies a coupling mechanism*: selection operates on behaviour, true belief is generally the cheapest route to adaptive behaviour, and the theory itself explains why the channel exists. Epiphenomenalism supplies none and cannot, because causal inertness is what it asserts. The discriminating criterion is therefore: **does the theory, by its own lights, provide a channel from truthmaker to report?**

An internal audit, not an external measurement. It also retroactively explains the illusionism stalemate — illusionism does supply a channel, and passes honestly rather than by evasion.

CORRECTION

I called this family transcendental, reaching for Kant. Overclaimed. Kant's arguments are constructive; this one builds nothing and only shows certain theories cannot be put forward. The right home is **Moore**: "P, but I don't believe P" is consistent and unassertable. The audit generates theoretical Moore-paradoxes — *epiphenomenalism is true, and my belief that it is was produced by processes to which its truth is irrelevant*. Consistent. Unassertable.

Stroud's 1968 objection to transcendental arguments lands squarely: they show at most what we must believe, never what is true, absent verificationism. That is verbatim the limit already flagged — binds belief, not reality. Which is quiet evidence for the classification: the thing fails in exactly the way transcendental arguments fail.

THE CAUSATION THREAT, AND WHY IT MISSES

If the criterion needs difference-making, and difference-making is interventionist, then supervenience appears to make it undefined: no intervention wiggles the phenomenal while holding the physical fixed, and Baumgartner argues this renders supervenient properties causally inert.

The argument proves too much, and the overreach is diagnostic. Temperature supervenes on molecular motion; solidity on lattice structure; being a recession on transactions. None survives the same test. The conclusion is not that consciousness is peculiar but that strict cross-level interventionism eliminates all higher-level causation — Kim's exclusion argument by another road, inheriting its familiar problem. Woodward's framework blocks it by design: causal claims are relative to a variable set whose members must be independent of one another, and putting a supervenient property and its full base in one set violates that condition before the machinery starts.

CORRECTION

But the better answer is that *we never needed to intervene on qualia*. The audit does not require establishing that consciousness is a difference-maker. It requires only the conditional: if a theory asserts the truthmaker is inert, then by its own lights there is no channel. Epiphenomenalism volunteers that. I reintroduced the external-measurement framing when posing this threat, contradicting the internal-audit result established one step earlier. The framing generated the difficulty.

THE ONE ESCAPE, AND WHAT IT COSTS

The epiphenomenalist can trade causation for **constitution**: Chalmers's direct phenomenal concepts, where the quality is not the cause of the belief but a constituent of it. Then the counterfactual fails for a non-causal reason — without the quale it would not be that belief. Acquaintance rather than a channel, and the undercutter is defeated with no arrow at all.

Take it seriously, and notice the cost. If the quale is a constituent of the belief and the belief causes the utterance, the quale is part of a cause. That is causal relevance in every sense that matters, and precisely the inertness the position was defined by. The escape works by ceasing to be epiphenomenalism, sliding toward constitution views, type-B physicalism, or the Russellian monism of §I.

Nothing has been proven false. A position has been shown to have no stable interior — keep the inertness and be unassertable, or add constitution and forfeit the inertness.

SPECIFICATION OF THE INSTRUMENT

TYPE	Self-applying assertability audit. Undercutting defeater with a self-referential trigger.
INPUT	A theory's own commitments. Nothing external, nothing measured.
TEST	Does T leave the truthmaker of a belief probabilistically inert with respect to that belief's production — $P(E \mid T) = P(E \mid \neg T)$? A noisy channel is not a cut one.
OUTPUT	Moore-paradoxical theories: consistent, unassertable.
IMMUNE TO	The exclusion problem. It never measures, so interventionist trouble with supervenience cannot reach it.
BINDS	Assertion, not truth. Stroud's ceiling.
CANNOT	Adjudicate rivals that both claim a channel. It will never settle physicalism against monism.
SCOPE	Filters explicit self-defeat. That is all it does.

A very small thing to have built from a thought about the moon. But it is an actual instrument with known limits, standing where there was previously only an intuition that theories eating their own evidence are somehow out. Most of what philosophy yields looks like this: the vertigo shrinks to a tool, the tool is narrower than the vertigo promised, and the narrowing is the result.

VII · RESULT THREE

Constitution is reference without information

The escape hatch from §VI, tested. It is a real relation. It secures almost nothing, and does not reach the case it was recruited for.

Load-bearing in three places. This section is numbered as one result among eight, which understates it. §V's mechanism depends on it, §VIII's case against the datum depends on it, and §VI's likelihood ratio breaks exactly where this section says the channel breaks. If constitution turns out to deliver content after all, three results move.

THE OBVIOUS OBJECTION FAILS

If the quale is a *part* of the belief, there are not two things standing in a relation but one thing overlapping itself — and containment is not an epistemic relation. Walls do not know their bricks. But this assumes constitution is doing the epistemic work by itself. It isn't. The claim is that constitution solves the **reference** problem, not the justification problem: the concept *this quality* picks out that quality because it has it as a part, and justification is then trivial precisely because reference is guaranteed. You cannot be wrong about which thing you meant, and cannot possess the concept in the thing's absence. Walls have no concepts. Containment is claimed as necessary, not sufficient.

THE REAL PROBLEM: DIRECT PHENOMENAL CONCEPTS CANNOT OUTLIVE THEIR QUALIA

If the quality is a constituent, the concept exists only while the quality does. Which rules out memory — *that red I saw yesterday* cannot be a direct phenomenal concept, the constituent is gone; generalisation — *red experiences are like this* needs a concept spanning tokens; and theory — every philosophical claim ever made about consciousness, epiphenomenalism included.

Recall what the hatch was for. The epiphenomenalist needed their belief *that epiphenomenalism is true* connected to phenomenal facts. That is a theoretical belief: formed over time, argued for, held while attending to no particular quale. It cannot be a direct phenomenal concept. Constitution might secure *I am having this*, occurrently. It cannot secure *there exist phenomenal properties and they are causally inert*.

The repair — the theoretical belief is inferred from occurrent direct ones and inherits their connection — fails exactly where predicted. Inference is a causal process. In the transition from *I'm having this* to *therefore qualia exist*, the work is done by the physical vehicle of the direct belief, not by its phenomenal constituent, which by hypothesis contributes nothing to anything. The channel breaks at the first inferential step, and everything downstream is a pure causal product to which phenomenality was irrelevant. **The audit's teeth survive a fair fight.**

THE GEOGRAPHY THIS REVEALS

Constitution is available only for occurrent, demonstrative, present-tense judgments. So the real epistemic split is not inner versus outer. It is **occurrent versus everything else** — acquaintance for the razor edge of attending to an experience now, description for all the rest, including all the rest of one's own mind. Nisbett and Wilson (1977) showed introspective reports of *why* are routinely confabulated. Schwitzgebel is the more uncomfortable finding, because it targets reports of *current* experience and finds people confident, inconsistent and wrong.

Which thins the edge further. If direct phenomenal concepts guarantee accuracy, why is introspection demonstrably bad? Only one answer is available: the guarantee covers the bare demonstrative and nothing else. You cannot be wrong that you are having *this*. You can be wrong about essentially anything you would go on to say about it.

There is a formal result in the neighbourhood, and it should be stated with its scope. Fields, Glazebrook and Levin prove that "adding observational, representational, or control capabilities to a meta-level component of a system cannot, even in principle, lead to a complete meta-level representation of the system as a whole," concluding that self-models "cannot, in general, be empirically tested by the systems that implement them." Breuer's earlier result is of the same family: no observer can distinguish all states of a system in which it is properly contained. Neither concerns phenomenal consciousness — Fields and colleagues extend explicitly to self-awareness, introspection and self-knowledge, and say nothing about qualia — so they do not settle anything here. What they supply is a formal backbone for what Schwitzgebel establishes empirically: the unreliability of introspection is not a contingent defect of human brains but a constraint on self-representation as such.

IS REFERENCE-WITHOUT-CONTENT DISTINCTIVE?

The profile is shared, and the deflation does not follow. Kripkean and Millian reference is fixed *externally*, by a causal-historical chain the speaker need not know; maximal ignorance about the referent is compatible with successful reference,

which is the entire point, and no infallibility whatever is delivered — you can be wrong about everything concerning Hesperus. The relation here is fixed from the subject's side, and what the absence of content buys is an accuracy guarantee rather than rigidity: infallibility about exactly one thing and nothing else. Same absence, incomparable payoffs.

The genuine precedent is nearer, and already in these references. Loar's phenomenal concepts are recognitional and directly referential with no descriptive mode of presentation — the same profile in the same domain. The difference is what it was built for: Loar's machinery defends type-B physicalism by explaining the epistemic gap without an ontological one. Here it is run against the acquaintance theorist instead. Same structure, opposite direction.

A real relation, securing a pointing with no content — too thin to perform any of the jobs it was recruited for.

WHAT IT REPAYS TO RESULT ONE

§V established that phenomenal facts are not selection facts but never said what they are instead. The mechanism on offer: selection requires a domain of discriminable candidates, and at the point where phenomenality is directly given there is no domain — just a demonstrative with no classificatory structure. The domain is built afterwards, by descriptive work.

TENSION · UNRESOLVED

This sits badly with §VIII. If phenomenality is *added* by classification, and classification is the fallible zone, then the "moment of givenness" is either not yet phenomenal — in which case the mechanism has no subject matter — or already phenomenal, in which case §VIII's case against the datum fails. The mechanism as first stated appeals to a pre-classificatory given while also holding the given has no content until described. It cannot have both.

The repair, and it changes the claim. Whether something is a selection fact is settled by the *structure of the situation*, not by the character of the given. Whether a domain of candidates is available is a fact about the architecture of the system, third-personally accessible, and knowable without knowing anything about the given's intrinsic character. So the mechanism should be stated architecturally: **indexical judgments presuppose their domain; phenomenal judgments constitute theirs.** That distinction survives §VIII intact, because it never appeals to the given having character.

And there is a fallback if the repair fails. Result One can rest on SLAM alone — but on SLAM *as licensed by the trilemma* in §V, not on the bare case. Bare SLAM is refusable by anyone willing to say the robot has minimal experience; what makes the refusal costly is the trilemma, which charges the objector wherever they stand. The mechanism here and the trilemma there do the same job by different routes. If the repair above fails, Result One stands on the trilemma and the mechanism goes.

VIII · THE REGROUND

The hard problem does not rest where it was thought to rest

THE DATUM TAKES REAL DAMAGE

To get from the bare demonstrative to *there is something it is like*, one must recognise the referent as having phenomenal character and then generalise into a claim about a kind. Both operations sit outside the guarantee. The pointing does not say what it points at — *this* is compatible with the referent being a functional state, a representational vehicle, anything.

Phenomenality is added by classification, and classification is the fallible zone. Anyone whose confidence rests on the sheer undeniability of the datum has lost that ground, and illusionism now explains both the belief and its error.

The standard rescue — you can be wrong about *what* the appearance is like but not that there *is* an appearance, since being wrong is itself a way of appearing — is where the debate has always deadlocked. Frankish's reply is that "seeming" equivocates between a functional state of representing-as, which illusionism needs and grants, and a phenomenally conscious appearance, which is what is in dispute. Worth adding: that reply is more radical than its name. If nothing appears there is no illusion, only a false belief with no experiential character. Illusionism is eliminativism about seeming. Conceding that does not refute it.

ANTICIPATED · KAMMERER (2022)

The conclusion of this section is not new. Kammerer argued that the Moorean argument against illusionism "fails (or is dialectically irrelevant)" because its defenders "cannot maintain that its crucial premise (properly understood) has the kind of support needed for the argument to work, without begging the question against illusionism." What is added here is the route. His is higher-order and epistemological — whether the realist can establish that their confidence has the standing required. This one is not: §VII established that the acquaintance guarantee covers a bare demonstrative with no classificatory content, from which it follows directly that *there is something it is like* is already a classification and already in the fallible zone. Same destination, by a road that does not require the realist to accept a higher-order constraint. This is the paper's thinnest novelty claim, and better said here than by a referee.

TOOL CHECK

Re-run the §VI audit on illusionism under the sharper question. Belief: *there is no phenomenality*. Truthmaker: the absence of phenomenality. The belief is caused by theoretical reasoning about physical facts, which are those truthmakers. It passes, exactly as before. The instrument returns the same verdict under pressure.

WHAT SURVIVES, AND WHERE IT WAS STANDING ALL ALONG

The hard problem has been thought to draw its force from the certainty of the datum. It does not. It draws it from §I — and §I never mentioned introspection. Newman's result is that pure structure is nearly vacuous: a structural description reduces to a cardinality claim plus mathematics. A world that is exhaustively structural is barely a world, because the relations have nothing to be relations *of*. Russell's demand for intrinsic natures was made in 1927 on purely logical grounds. Ontic structural realists bite the bullet; most take it as a reductio; either way the demand is generated by the form of physical theory, not by anyone's confidence about their inner life.

Physics is exhaustively structural.
Pure structure is vacuous. (Newman 1928)
∴ Something non-structural obtains.
∴ There are facts physics does not reach.

No step passes through the datum.

But be precise about the delivery. This establishes a gap for *intrinsic natures*, not specifically for phenomenal ones. The move from *intrinsic natures exist* to *experience is among them* still runs through the datum. So the hard problem survives as a conditional whose antecedent has been substantially weakened and whose consequent has been strengthened and made independent of introspection entirely.

What was demolished is the epistemology. What is standing — Leibniz's mill, structure and dynamics, Newman — is all metaphysics, and none of it depends on how sure anybody feels.

A check that the reasoning is not overshooting: the cogito survives by these lights at exactly the thinness identified — *there is thinking* is near enough the bare demonstrative — and no further. Descartes's inference to a thinking *thing* with phenomenal character overshoots, which is Lichtenberg's objection that the honest form is *it thinks*. The result reproduces a known boundary rather than inventing one.

IX · RESULT FOUR

Newman against Φ

Question 02, run. The result is a dilemma that generalises past IIT to any theory answering §I with a formal quantity. Kleiner generalises the *problem* — Newman against structural assumptions throughout philosophy and science; what generalises here is the *fork*, which is a claim about what the problem costs.

THE THEORY, STATED TO BE ATTACKED

Five axioms about experience — intrinsicity, composition, information, integration, exclusion — converted into postulates on a physical substrate, then an *identity*: an experience is the maximally irreducible cause-effect structure unfolded from a substrate. Not correlated with. Identical to. Φ is computed over the **perturbational** transition probability matrix — not the states a system visits but its response to every possible state of its units, partitioned every way, with the minimum information partition fixing irreducibility and the maximum of Φ fixing which overlapping candidate wins.

IIT IS NOT NAIVELY VULNERABLE

Newman bites claims of the form *all we know is that reality has structure S*, because relations are extensionally specified and nothing pins which ones really obtain. IIT pins them: the real relations are causal ones, read off actual cause-effect power. That is the work intrinsicity does in the axiom list, and any attack ignoring it is hitting a strawman.

So push on the cash value. Cause-effect power is measured interventionally — the perturbational TPM is a matrix of conditional dependencies under forced settings. And a matrix of conditional dependencies is a structure. The criterion meant to lift IIT clear of structuralism turns out to select *which* structure to compute, over which units, at which grain. Intrinsicity names a privileged structure; it does not name something other than structure.

THE PROBLEM IS NOT NEW, AND THE PUBLISHED REPAIR IS INSTRUCTIVE

Kleiner put Newman to consciousness science with IIT as the central case and reached the conclusion the last two paragraphs reach: "IIT does not provide a phenomenal interpretation of this structure, it only provides the mathematics," so its structural claim is "entirely void, over and above the cardinality of the elements in the structure." Priority is his, and so is the connection to what Chalmers calls the Rosetta Stone problem — if Newman applies, there is no translation from the mathematics into phenomenology waiting to be found. What is left open is what the problem costs a theory unwilling to give up the identity. That is this section, and the answer is sharpened rather than weakened by his repair, which is the strongest on offer.

The repair replaces the bare existence condition on relations with a phenomenal one. A phenomenal property p is an **R-property** iff "any variation that preserves the relation R preserves p ," and the structural claim is true only if "the relation R as specified by E exists, and there is a corresponding higher-order phenomenal R -property." Newman is genuinely blocked, and blocked for the right reason: no phenomenal property can be conjured by stipulating relations on a set of the right cardinality. Nothing in the theorem delivers one.

But notice what does the blocking. Content is supplied by phenomenal facts at both levels — phenomenal properties as the elements, higher-order phenomenal properties as the individuator of the relations. Which is fatal to the one thing IIT was relying on. **Intrinsicality was supposed to pin the relations by cause-effect power; under the repair they are pinned by phenomenal invariance instead.** A theory may have the repair, or it may have intrinsicality doing the work. Not both.

HORN ONE — NOTATIONAL INTRINSICALITY

Then Newman bites. Any set of adequate cardinality can be given relations instantiating the Φ -structure, and the identity claim carries no content beyond cardinality plus mathematics. This is not abstract: Aaronson found by explicit construction in 2014 what Newman guarantees by theorem — XOR grids and expander-graph systems with enormous Φ and nothing cognitively interesting occurring. Tononi bit the bullet and accepted that the grid is conscious. An available reply, but the reply of someone whose theory has been shown cheaply satisfiable, which is Newman's conclusion.

In fairness, a refinement. IIT can identify experience with a *token* concrete structure rather than the abstract type, and Newman does not straightforwardly bite token identities. That works — but everything IIT *predicts* lives at the type level: cerebellum low and cortex high, feedforward systems near zero, this anaesthetised patient unconscious. The token move rescues the bare identity assertion and leaves the entire predictive content exposed. Newman bites the part that does the work.

What the repair does to this horn. It answers it, and the price is the second horn. Ask why the higher-order phenomenal property carries exactly the invariance profile of the relation it certifies, and the theory is silent: the covariance is posited, not derived. That is the explanatory gap again, now itemised — one brute connection per relation where there was previously one brute identity. More posits, the same debt. And the restored content is not measurable content: R -properties are phenomenal, the physical dynamics is fixed by the structure alone, so by §XIII's criterion they make no difference to the states that would report them. The repair moves IIT from notationally void to contentful and permanently undetectable — off horn one and onto horn two.

One limitation of the repair does *not* help here, and it is worth recording that it was checked. Kleiner notes that relations with trivial automorphism groups cannot be told apart by the criterion, which would matter if Φ -structures over real substrates were generically rigid. But his own conditions exclude the degenerate case — he argues "there cannot be an R -property for a relation that has trivial automorphism" — and he is explicit that where a structure contains more than one relation and the group is trivial only collectively, "all is well." A Φ -structure is exactly that case. The objection does not bite.

HORN TWO – METAPHYSICAL INTRINSICALITY

Suppose instead cause-effect power is a genuine primitive, not analysable into counterfactual patterns. Then Newman is blocked, and blocked properly: a set of abstract objects has no powers, so the relations cannot be freely stipulated the way relations on a bare set can. This is the known escape route — Chakravartty's causal properties, dispositional essentialism — and IIT is entitled to it.

FALSIFIED IN REVIEW

First pass claimed horn two collapses IIT into Russellian monism. Wrong, and the correction is more damaging. Russellian monism locates the phenomenal in the *relata*; IIT locates it in the *relations*, with powers as non-phenomenal primitives. So it is a powers-based structuralism, not a monism — and the real cost is elsewhere.

The cost is this. If experience is the organisation of non-phenomenal powers, IIT owes an account of why that organisation is *experience* rather than merely organisation. It does not supply one. The axioms are taken as self-evident and the identity is postulated. **The identity claim is a brute posit connecting non-phenomenal primitives to phenomenal facts — which is the explanatory gap it was meant to close, converted into an axiom.** And the unfolding margin stands independently: any recurrent system has a feedforward equivalent with identical behaviour and $\Phi \approx 0$, so the theory is adjudicated against intuitions about its axioms rather than against measurements. §I's missing observable, arriving on schedule.

THE DILEMMA, AND WHY IT IS NOT ABOUT IIT

NOTATIONAL intrinsicality → Newman consumes the predictive content.
METAPHYSICAL intrinsicality → Newman blocked, but the identity is a brute posit, the gap is axiomatised, and an unfalsifiable margin remains.

PHENOMENAL individuation (Kleiner's repair) → Newman blocked and content restored – by phenomenal posits, which is the second horn with the primitive relocated. A route between the horns, not an exit.

It stays a dilemma. The third line is a path, not a third option, and promoting it to one would be the same overclaim this section convicts IIT of.

Any theory answering §I's demand with a formal quantity meets the same fork, because formalisation delivers structure by construction and §I established that structure cannot be the whole story. This explains something usually blamed on failure of imagination — the sense that every mathematical theory of consciousness changes the subject. It is a consequence of what formalisation is.

SECOND INSTRUMENT

"Declare your primitive" is trivial as a test — every theory has something basic. The version with teeth locates *where the phenomenal enters*, and there are only three places. It is primitive (panpsychism, Russellian monism): honest, and pushes everything onto combination. It is denied (illusionism): consistent, and owes an account of the reports, which it gives. Or it is derived from non-phenomenal primitives — in which case the derivation is either real, and should be shown, or a posited identity, and the gap is being axiomatised rather than closed.

Formal theories systematically claim the third and deliver a disguised posit, because formalism manufactures the appearance of derivation.

Structurally this is the same kind of instrument as §VI: read the theory's own commitments, locate one specific feature, check whether it is present or disguised. No external measurement, and the same ceiling — IIT is not shown false, only shown not to be doing what it says it is doing.

X · RESULT FIVE

Combination is not a physics problem — and a physics argument proves it

Question 06, run, beginning with whether we are entitled to the tool. We are not, and the reason is the result.

THE ENTITLEMENT CHECK

Renormalisation operates on a space of theories parameterised by couplings, flowing under coarse-graining and rescaling. For proto-phenomenal properties to be subject to it they would have to appear as operators with couplings in an effective action. So the entitlement question is precise: **does panpsychism give the phenomenal a coupling?**

Russellian monism explicitly does not. Its whole content is that proto-phenomenal properties are the intrinsic natures *of the relata already there* — not additional properties with their own dynamics. That is why it is not dualism and why it is compatible with the causal closure of physics. The phenomenal is given no term in the action by design.

RG cannot see the phenomenal, not because the tool is unsuited but because the position deliberately declines to give it anything the tool operates on.

PREDICTION KILLED

The previous turn predicted a dilemma — proto-phenomenal properties are either irrelevant operators and idle, or relevant and empirically visible-but-absent. Malformed: it presupposed exactly the entitlement the check denies. They are not operators at all. The check earned its place by destroying the result it was run to protect.

WHAT SURVIVES IS A FORCING MOVE

The dilemma does apply to any panpsychism that *does* give the phenomenal independent physical influence. There the fork is lethal: a relevant coupling belongs in the low-energy effective description of brains and of everything else at that scale, and no such term is missing from any theory that works; an irrelevant one flows to zero and contributes nothing to macro experience. So panpsychism is forced into the Russellian form — and the Russellian form is precisely the one that cannot use physical composition to solve combination.

Panpsychism with a coupling → dilemma: idle, or visible and absent.

Panpsychism without a coupling → invisible to every compositional tool physics has.

The inapplicability is forced, not accidental.

SO WHAT IS COMBINATION A QUESTION ABOUT?

Composition of intrinsic natures — which have no dynamics, so nothing dynamical can govern how they compose. James's original objection stands undisturbed by any amount of physics: a hundred feelings do not make one feeling of a hundred. Goff's subject-summing problem is the sharper modern form — even granting quality combination, subjects do not sum, and it is unclear why a macro subject should exist at all rather than many micro ones.

Run the §IX instrument across the live proposals and it sorts them cleanly. **Phenomenal bonding** posits a relation binding micro-experiences into macro ones whose nature is unknown — a posit presented as a solution, failing the test exactly as IIT's identity claim does. **Cosmopsychism** reverses the direction: the cosmos is the fundamental subject and individual subjects are derived by decomposition. That is not a disguised posit but an honest change of which fact is fundamental, and it passes the instrument — at the cost of owing a decombination problem instead.

THE ONE CONTRIBUTION PHYSICS DOES MAKE, AND IT IS NEGATIVE

Universality: systems with wildly different microphysics flow to the same fixed point and share critical exponents, because macro behaviour is insensitive to almost all micro detail. Apply that to the standard motivation for panpsychism — the genetic argument, that experience cannot emerge from wholly non-experiential ingredients, so the ingredients must be experiential. Emergence of macro properties absent at the micro level is not merely possible but routine, and RG explains exactly why the micro details wash out.

Strawson's reply is available and correct: those are all *structural* macro properties, and structure can be built from structure, whereas experience is non-structural. Which concedes the point in the form that matters. The genetic argument then contributes nothing beyond §I's claim that experience is non-structural — a claim already established there, on Newman's grounds, without any appeal to intuitions about emergence. **Universality does not refute the genetic argument; it strips it of independent force.**

XI · METHOD NOTE

Both instruments are filters

Every result to §X has been a narrowing. The vertigo shrank to a tool; inner reality to a razor edge; the hard problem's datum to a pointing; IIT to a posit; combination out of physics entirely. The explanation is unflattering and structural: the assertability audit detects self-defeat, the second instrument detects disguised posits, and **neither can construct anything**. We own a demolition kit, so everything we touch gets smaller.

Five consecutive subtractions is evidence about the method, not about the world.

§XII was taken partly as a corrective — a question where the answer could come back positive. It did. And §XIII supplies the missing kind of instrument outright: inverting the coupling criterion turns a filter into a test that sorts positions rather than only eliminating them.

ON THE APPARATUS ITSELF

A reasonable objection to this document is that numbered results, named instruments and a cleared register are scientific dress on what is straightforwardly philosophical argument. The vocabulary is not a claim to empirical status. It is bookkeeping, and it earns its place by making retraction cheap and visible: §V's mechanism is flagged as contested, §XIII's closure of the acquaintance exit is marked as overstated and narrowed, §X records a prediction destroyed by its own entitlement check, §IX records a claim falsified in review, §VI records a defining clause falsified by Elga's counterexample, and §XI is this admission. A conventional essay absorbs its revisions into the prose and shows the reader only the surviving version.

"Instrument" is doing real work too. It marks a reusable argumentative test with a stated scope and known failure modes, which is a category philosophy already has — Occam's razor, the open question argument, transcendental arguments — and naming it commits you to specifying the scope, which §VI and §XIII both do.

The charge would bite if the frame made the arguments look stronger than they are. It has done the reverse: it is what forced this document to disclose its own bias, and to keep the wreckage of five withdrawn claims on the page.

XII · RESULT SIX

The gradient fixes the direction of felt time, not its texture

Question 04, run. The first result here that is mostly constructive: four components grounded, one residue isolated, and the residue is the one whose existence is itself disputed.

THE EXPLANANDUM IS FIVE THINGS, AND THE BUNDLING IS WHY THE DEBATE STALLS

- a direction past fixed, future open
- b memory records of earlier and not later
- c duration the specious present; experience has extent
- d flow the sense that time *moves*
- e presentness this moment is privileged

DIRECTION REDUCES — AND BY A STRONGER ARGUMENT THAN THE STANDARD ONE

The usual route is Albert's: records require a ready state, ready states require the Past Hypothesis, so retrodiction is only reliable because the universe began in low entropy. Solid, but it establishes dependence, not identity.

The inversion case does better. Consider an observer in a region where entropy decreases toward what we call the future. Their records form toward their own low-entropy end, so they experience as past what we experience as future. Now ask which of you is right about the direction of passage. **There is no such fact.** Both of you experience time as running toward increasing entropy; you use "future" for opposite directions of the manifold and disagree about nothing. The comparative question is not truth-apt.

If felt direction were a fact beyond the gradient, the comparative question would be substantive. It isn't. So felt direction is the gradient, experienced from inside.

Note the argument does *not* generalise to flow: both observers have flow. It reduces the direction and leaves the flowing untouched, which is the first evidence that the components come apart.

PRESENTNESS IS DISPATCHED BY §V

Now is a paradigm indexical, so presentness is a selection fact — it picks a moment from a domain already given. §V established that indexicality is cheap and requires nothing phenomenal: any system with a clock and a self-model has a "now," as any system with SLAM has a "here." Presentness therefore leaves the mystery pile entirely. It never belonged there.

DURATION IS NEURAL, NOT THERMODYNAMIC

The specious present has an extent — you hear a melody, not a sequence of instantaneous tones — and its magnitude is set by neural integration windows, on the order of seconds, with the exact figure contested. Thermodynamically constrained, certainly; thermodynamically determined, no. It is a fact about brains. And Husserl's point matters here: retention is not recollection. The just-past tone is given *as present-and-fading*, not re-presented as past, so having records does not explain the structure. Component c is untouched by the entropy story in either direction.

THE RESIDUE

Flow is left, and the debate over it is a genuine stalemate. Prosser's argument is the strongest against: if time did not pass, experience would be exactly as it is, so experience is not evidence for passage. The inflationist cannot say what passage would add — the rate question is fatal, since time passing at one second per second is a ratio of one, not a rate. Neither party can specify what would settle it.

Parasitic, not identical — and the parasitism is component-specific. Direction reduces to the gradient. Memory reduces to records. Presentness is indexical. Duration is neural. Only flow survives unexplained, and its reality is what the surviving dispute is about.

A PATTERN, THIRD SIGHTING

Prosser's argument has a shape we have now met three times. §I: no observable distinguishes a conscious system from a functionally identical unconscious one. §IX: no behaviour distinguishes a recurrent system from its unfolded feedforward equivalent. §XII: no experience distinguishes a passing world from a static one. Three independent literatures, one structure — *no possible evidence separates X from not-X*. That is no longer a coincidence.

XIII · RESULT SEVEN

One phenomenon, and we already owned the tool for it

Question 05, run. The three undetectabilities are one, the one is §VI's coupling criterion arriving from three directions, and inverting it yields this investigation's first constructive instrument.

TWO OF THE THREE COLLAPSE INTO THE FIRST

The unfolding argument is not independent. A recurrent network and its unfolded equivalent are perfectly distinguishable physically — you can look at the wiring. Only the *consciousness* difference is invisible, and it is invisible because report is the whole evidence base for consciousness. Which is true only because consciousness has no observable of its own. §IX's undetectability is a consequence of §I's, not a second instance.

Prosser's argument is our own criterion, independently derived. His reasoning is that to detect X a state must be counterfactually sensitive to X; experiential states are fixed by physical states; if passage is a further fact not fixed by the physical, experience cannot covary with it. That is §VI's audit applied to passage, arrived at from the metaphysics of time by someone not doing philosophy of mind.

And §I's own case is the same: a zombie's absent qualia make no difference to its reports, which is the paradox of phenomenal judgment. Three literatures, one criterion.

A putative fact *F* is undetectable exactly when *F* makes no difference to the states that would report it.

Which is the §VI coupling criterion, stated as a claim about evidence rather than as a claim about assertability.

WHAT THE IDENTIFICATION BUYS

First, a demotion. Each literature treats its undetectability as a deep puzzle about its own subject matter — consciousness mysteriously beyond measurement, IIT mysteriously untestable, passage mysteriously imperceptible. On the unified reading none of these is about the subject matter at all. Each is a report that *the posited fact has been given no causal work to do*. That is a fact about the posit.

Second, and more usefully: **invert the criterion and it stops being a filter**. Instead of asking whether a theory supplies a channel, ask what a fact must be like to be detectable at all — it must make a difference to the states that report it. That sorts theories of consciousness into those that could in principle become empirical and those that could not, ever.

POSITION	SCIENCE-ELIGIBLE	WHY
Illusionism	Yes	The representational states are themselves the difference-makers, and they are detectable.
Type-B physicalism	Yes	The phenomenal is a physical property; detect that property and you have detected it. Still owes the explanatory gap.
IIT (horn two)	Partial	The organisation is detectable. Its being <i>experience</i> is not.
Russellian monism	No	Intrinsic natures are the categorical bases of causal powers — swap them and the structure is unchanged, so nothing differential is contributed.
Epiphenomenalism	No	No channel by definition. Permanently outside measurement.

THE STING

Notice which positions pass: exactly those that either deny the phenomenal or identify it with something physical. Every position that preserves it as something further fails. That is not an argument that any of them is false. It is an argument about what follows if one is true.

The missing observable is not a temporary embarrassment awaiting better instruments. It is a disjunction — either consciousness is not what the non-physicalists say it is, or there will never be a science of it — and no evidence can decide which, by the criterion in question.

THE ESCAPE, AND HOW FAR IT IS ACTUALLY CLOSED

All of this assumes detection requires counterfactual sensitivity, which is an externalist assumption about evidence — and precisely what the acquaintance theorist denies. Their claim is that phenomenal facts are known by containment rather than by causal sensitivity. So the constitution route is the one exit from the dilemma.

An earlier version closed this exit by citing §VII, on the ground that constitution delivers reference without information. That is a result about *content*; the externalist assumption is a thesis about *evidence*. One can accept §VII entire and still deny that detection requires covariation. The exit was declared closed more firmly than the argument earned.

What can be said is narrower. Detection is discrimination — telling one state of affairs from another — and discrimination requires content, which by §VII the acquaintance relation does not supply. So even granting acquaintance as a genuine non-causal mode of access, it cannot perform the function the exit needs: separating two hypotheses.

The acquaintance theorist has a reply, and it is coherent. Acquaintance is not discrimination but *presence* — you do not tell that phenomenality obtains, you have it, and demanding discrimination begs the question against them. Nothing here refutes that. What follows is only this: **a mode of access that is neither discriminative nor describable cannot function as evidence in a dispute between parties**, since settling a dispute requires something that could have come out otherwise and can be shown. Acquaintance-as-presence may well be a genuine epistemic relation. It is not an exit from an indiscriminability result.

Self-check with the §VI audit, which is owed here: is belief in *this* result coupled to its truthmaker? Yes — it is derived from the structure of the positions, not from observing consciousness. A conditional about what follows from what. It passes.

XIV · RESULT EIGHT

The two survivors cannot be separated

The last question, run. The criterion that selected the field is the same criterion that cannot discriminate within it — and this generalises.

HALF THE DIFFERENCE DISSOLVES ON INSPECTION

Frankish already draws the distinction that matters. **Weak illusionism** is illusionism about the problematic properties only — ineffability, intrinsicality, direct apprehension — and is compatible with realism about experience under a modest description. That is type-B physicalism in different clothes, and Frankish says as much. The genuine rival is **strong illusionism**, on which phenomenality itself is not instantiated.

WHAT REMAINS, STATED FAIRLY

Both positions agree that only physical properties are instantiated; that a special class of states — phenomenal concepts, or introspective representations — presents physical states in a distinctive way; that this presentation generates the impression of something irreducible; and that the impression is misleading. **The architecture is identical.** They differ on the diagnosis of a shared error: type-B says we have a misleading concept of a real property; strong illusionism says we have an accurate concept of a property that is not instantiated.

And the property in question is the one §VII showed cannot be characterised — the acquaintance zone secures reference with no content. So the dispute is an affirmation and a denial of the instantiation of something neither side can describe.

APPLY THE CRITERION

Both agree on the complete physical description, the complete functional and representational description, every utterance, every judgment, every behavioural disposition, and the existence and workings of the introspective mechanism. There is no

state whose occurrence differs between them.

A fact is detectable iff it makes a difference to the states that would report it.

Type-B and strong illusionism differ in no reporting state.

∴ The difference between them is undetectable.

Note this is consistent with §XIII rather than in tension with it. Science-eligibility asks whether a theory's own posits do causal work. Inter-theoretic discriminability asks whether two theories differ in their posits' effects. A theory can pass the first while its rivalry with another fails the second. And by the standing Stroud ceiling, undetectable does not mean meaningless — no evidence bears on the question, which is not the same as the question having no answer.

THE CONFIRMING INSTANCE

The obvious objection is that theories of consciousness *are* tested against each other — the COGITATE consortium ran IIT against global neuronal workspace theory with preregistered, competing predictions. That is the right challenge, and the experiment turns out to illustrate the result rather than threaten it.

The predictions concerned where and when stimulus attributes are neurally represented and interareal synchrony. No Φ was computed and global broadcast was not measured directly; consciousness was not assessed trial by trial but assumed from stimulus conditions — suprathreshold, high-contrast, foveally presented. The consortium says why in its own words: "Both theories have a mathematical or computational core (integrated information for IIT and global workspace for GNWT) and proposed biological implementations," and "Although it is difficult to test the mathematical or computational core of either theory directly, their competing biological implementations are empirically testable with current methodologies." What was adjudicated was the bridging hypotheses — the place where the two theories genuinely disagree about physical facts.

And the consortium states the consequence itself: "One consequence of this biological focus is that theorists could respond to challenging data by modifying the proposed biological implementation while retaining the mathematical or computational core of a theory." That sentence is this section's result, written by the people running the experiment. Both sides duly said so afterwards — for GNWT, that because the study "only used clearly visible stimuli, it could not test the most central predictions of the GNWT"; from the IIT side, that the experiments "only tested some idiosyncratic predictions made by certain theorists, which are not really logically related to the core ideas of IIT." The consortium declined a verdict: "we invite readers to weigh the evidence themselves—".

The experiment was possible because the theories disagree about physics. It was inconclusive about the cores because the cores do not. §XIV's antecedent fails in this case, and that is the point rather than a limitation.

OWN ANALYSIS • NOT CONSENSUS

No published commentary frames the COGITATE predictions as testing access rather than phenomenality, and nobody has connected the unfolding argument or Kleiner and Hoel's falsification result to this experiment in print. The consortium cites Lakatos and Popper and none of that literature. The reading above is therefore an interpretation offered here, not a citable consensus.

THE GENERALISATION, WHICH IS THE REAL RESULT

Science-eligibility requires locating the phenomenal among the physical facts. Any two theories that do so while agreeing on all physical facts differ in no reporting state. Therefore:

Any two science-eligible theories of consciousness that agree on the physical facts are mutually indiscriminable. The property that admits a theory to the field is the property that prevents the field from being adjudicated.

Which puts §I's missing observable in a harder light than when we found it. A science of consciousness can tell you which physical states do what. It cannot tell you which description of those states is the correct one — not for want of instruments, but because the criterion that makes a theory eligible for measurement is the same criterion under which eligible theories coincide.

XV · NOT SOLVED

The taste of chocolate

Stated plainly, because a document with eight results and three instruments can read as more triumphant than it is. The hard problem is not solved here. It is not close to solved.

What happened is that the classic case turned out to be four questions with sharply different statuses.

THE QUESTION	STATUS	WHERE IT STANDS
Why chocolate tastes different from coffee	Answered	Receptor chemistry, discriminative structure, cortical representation. A structural question, and structure is what science does well.
Why you say it tastes like this	Answered	A causal chain from tongue to utterance. Fully tractable.
Why you cannot convey it	Settled here	§VII. The acquaintance zone gives reference with no content, and ineffability is what operating a contentless demonstrative feels like from inside. Nothing is being hidden; there was never content there to fail to transmit.
Why there is any tasting at all, rather than mere discriminating	Untouched	The hard problem, exactly where it was.

What changed around the fourth is that it no longer rests on your certainty that there is something it is like. It rests on Newman: physics is exhaustively structural, pure structure is vacuous, so something non-structural obtains. Firmer ground, delivering less. It gets you intrinsic natures. It does not get you chocolate — the bridge from *intrinsic natures exist to experience is among them* still runs through the datum, and §VIII showed the datum is a classification, which is where introspection demonstrably fails.

So the position is slightly worse than unsolved. The question stands, and the confidence that made it urgent has been undermined, without the question dissolving — because the Newman route survives independently of it.

One reading to resist above all others: **elimination is not progress toward an answer**. The field was narrowed to two positions and then the two were shown to be unadjudicable. That is not convergence on chocolate. It is a demonstration that converging is not the available motion.

XVI · CLOSE

Where the register ends

Every question on the original list has been run. Eight results and one method note, three instruments, and a field narrowed from the whole literature to two positions that cannot be told apart.

The thought this began from was that reality is rendered, that it is amazing anything is happening, and that whatever is perceived depends on retina and optic nerve and processing. The rendering intuition survived and sharpened into Newman's problem. The dependence claim survived and got worse — the zone of guaranteed access is a razor edge and the rest is fallible description. The wonder did not survive in the form it had; its datum moved from the secure zone to the fallible one, and what replaced it stands on the form of physical explanation rather than on anyone's certainty.

The mystery changed address twice: from the moon to the looking, and then from the looking to the form of physical explanation — which is in neither.

Two unforced threads remain open and neither is load-bearing. The composition question from §X, reframed: metaphysical grounding versus Husserlian moments, tested against subject-summing. And the §IX instrument has not been run across global workspace, higher-order theories, predictive processing or the attention schema, where the result is expected to be confirmatory.

APPARATUS

References

- Aaronson, S. — Why I am not an integrated information theorist. 2014. (Expander graphs with large Φ .)
- Albantakis, L. et al. — Integrated information theory (IIT) 4.0. *PLoS Computational Biology*, 2023.
- Armstrong, S. — Anthropic decision theory. 2011/2017.
- Baumgartner, M. — Interventionism and epiphenomenalism. *Canadian Journal of Philosophy*, 2010.
- Bong, K.-W. et al. — A strong no-go theorem on the Wigner's friend paradox. *Nature Physics*, 2020.
- Bostrom, N. — *Anthropic Bias*. 2002. (SSA, SSSA, the reference class problem.)
- Breuer, T. — The impossibility of accurate state self-measurements. *Philosophy of Science* 62(2), 1995, 197–214.
- Carroll, S. — Why Boltzmann brains are bad. 2020. (Origin of "cognitively unstable.")
- Chakravartty, A. — *A Metaphysics for Scientific Realism*. 2007. (Causal properties as the escape from Newman.)
- Chalmers, D. — *The Conscious Mind*, 1996, ch. 5 (paradox of phenomenal judgment); The content and epistemology of phenomenal belief, 2003; The Matrix as metaphysics, 2005.

Cogitate Consortium (Ferrante, O., Gorska-Klimowska, U. et al.) – Adversarial testing of global neuronal workspace and integrated information theories of consciousness. *Nature* 642(8066), 2025, 133–142. See also Naccache, L. et al., *Neuroscience of Consciousness* 2025(1): niaf037 (note published correction); and IIT-Concerned et al., *The Integrated Information Theory of Consciousness as Pseudoscience*, PsyArXiv 2023 (non-final version, targeting the 2023 preprint).

Demopoulos, W. & Friedman, M. – Critical notice: Bertrand Russell's *The Analysis of Matter*. *Philosophy of Science* 52(4), 1985, 621–639. Cf. Demopoulos, Three views of theoretical knowledge, *BJPS* 62(1), 2011.

Deutsch, D. – Quantum theory of probability and decisions. 1999.

Doerig, A. et al. – The unfolding argument. *Consciousness and Cognition*, 2019. (Reply: Kleiner & Hoel, *Falsification and consciousness*, 2021.)

Elga, A. – Boltzmann brains and cognitive instability. *Philosophy and Phenomenological Research*, 2025. (Labelscramble; the likelihood-ratio treatment.)

Esfeld, M. – Quantum entanglement and a metaphysics of relations. *SHPMP* 35(4), 2004, 601–617; The modal nature of structures in ontic structural realism. *ISPS* 23, 2009.

Fields, C., Glazebrook, J. F. & Levin, M. – Principled limitations on self-representation for generic physical systems. *Entropy* 26(3):194, 2024.

Frankish, K. – Illusionism as a theory of consciousness. *JCS*, 2016. (Weak vs strong illusionism.)

Frauchiger, D. & Renner, R. – Quantum theory cannot consistently describe the use of itself. *Nature Communications*, 2018.

Fuchs, Mermin & Schack – An introduction to QBism. 2014.

Goff, P. – *Consciousness and Fundamental Reality*. 2017. (Subject-summing; phenomenal bonding; cosmopsychism.)

Haldane, J. B. S. – *Possible Worlds*. 1927.

Helmholtz, H. von – *Handbuch der physiologischen Optik*. 1867. (Unconscious inference.)

Husserl, E. – *On the Phenomenology of the Consciousness of Internal Time*. (Retention is not recollection.) Cf. Dainton, *Stream of Consciousness*, 2000.

Ismael, J. – *The Situated Self*, 2007; *How Physics Makes Us Free*, 2016.

James, W. – *Principles of Psychology*. 1890. (The combination problem.)

Kammerer, F. – How can you be so sure? Illusionism and the obviousness of phenomenal consciousness. *Philosophical Studies*, 2022.

Kleiner, J. – The Newman problem of consciousness science. *Philosophy and the Mind Sciences*, special issue *Structuralism in the Science of Consciousness*. [vol./issue/year to confirm from the article page.] (Newman applied to consciousness science; IIT as central case; the R-property repair.) Cf. the companion piece in the same issue, *Structural representation and the Newman problem of the brain and AI*.

Kriegel, U. – *Subjective Consciousness*. 2009. (Self-representationalism.)

Kripke, S. – *Naming and Necessity*. 1980. (Reference without descriptive content, fixed externally.)

Ladyman, J. – What is structural realism? *SHPS A* 29(3), 1998, 409–424. Ladyman, J. & Ross, D. – *Every Thing Must Go*, 2007. (Ontic structural realism; the Newman escape at p. 128.)

Leibniz, G. W. – *Monadology* §17. 1714. (The mill.)

Lewis, D. – Attitudes de dicto and de se, 1979 (the two gods); What experience teaches, 1988.

Lichtenberg, G. C. – *Sudelbücher*. (Against the cogito: *it thinks*, not *I think*.)

Loar, B. – Phenomenal states. 1990/1997. Cf. Papineau, *Thinking About Consciousness*, 2002; Block & Stalnaker, 1999. (Type-B; the phenomenal concepts strategy – and Chalmers's dilemma for it, 2007.)

Neal, R. – Puzzles of anthropic reasoning resolved using full non-indexical conditioning. 2006.

von Neumann, J. – *Mathematische Grundlagen der Quantenmechanik*. 1932.

Newman, M. H. A. – Mr. Russell's causal theory of perception. *Mind*, 1928.

Nisbett, R. & Wilson, T. – Telling more than we can know. *Psychological Review*, 1977.

Perry, J. – The problem of the essential indexical. *Noûs*, 1979.

Plantinga, A. – *Warrant and Proper Function*, 1993. (EAAN. Critique: Fitelson & Sober, 1998.)

Pollock, J. – *Contemporary Theories of Knowledge*. 1986. (Undercutting vs rebutting defeaters.)

Prosser, S. – *Experiencing Time*. 2016. (Passage is not perceivable.)

Russell, B. – Knowledge by acquaintance and knowledge by description, 1910; *The Analysis of Matter*, 1927.

Schaffer, J. – Monism: the priority of the whole. *Philosophical Review*, 2010.

Schwitzgebel, E. – *Perplexities of Consciousness*. 2011. (Introspection of current experience is unreliable.)

Sebens, C. & Carroll, S. – Self-locating uncertainty and the origin of probability in Everettian quantum mechanics. *BJPS*.

Strawson, G. – Realistic monism. *JCS*, 2006. (The genetic argument; cf. Nagel, Panpsychism, 1979.)

Stroud, B. – Transcendental arguments. *Journal of Philosophy*, 1968.

Tononi, G. & Koch, C. – Consciousness: here, there and everywhere? *Phil. Trans. R. Soc. B*, 2015.

Wallace, D. – *The Emergent Multiverse*. 2012.

Woodward, J. – *Making Things Happen*, 2003; Interventionism and causal exclusion, 2015.

Zeh, H. D. – On the interpretation of measurement in quantum theory. 1970. Zurek, W. – Einselection, 1981–82.