

Epistemic Horizons: A Computational Theory of Consciousness, Perspective, and Knowability

Jason K. Resch

Abstract

Computational functionalism leaves open what fixes the character of experience, which process possesses it, how it causes behavior, and how a subject can know it. This paper defends the centered-knowledge thesis: for a phenomenal center, phenomenal character is identical to the complete organization of the distinctions available at that causal locus, their access relations, and the transitions they can govern. Because the same centered states constitute phenomenal differences and produce comparison, memory, judgment, report, and action, the thesis explains epistemic contact with one's experience, answers the causal demand of the meta-problem, excludes functionally identical zombies, and explains the explanatory gap through two consequences of occupation: to occupy one complete state is to occupy no other, and to occupy a state is not to be shown its realization. Organizational invariance, substrate independence, and multiple realizability follow from the identity rather than entering as separate principles, and the substitution level turns them into a criterion for replacement, emulation, and upload fidelity. Causal access individuates the subject, explains binding, unity, and privacy, and permits nested and overlapping centers. The same boundary that individuates a subject also limits what it can know: present centered knowledge underdetermines history, program, substrate, and world, and physics appears to any observer as structure and dynamics because only these cross a point of view's boundary. The identity's natural limit, advanced as a separable hypothesis, is that phenomenality begins with the first causally usable distinction. Comparisons with global-workspace, higher-order, and integrated-information theories yield contrasting attributions, and tests are proposed. Experience is what structured knowledge is like for the process that possesses it.

Keywords: consciousness, computational functionalism, centered knowledge, explanatory gap, organizational invariance, artificial consciousness

1 Introduction

1.1 Follow the difference

A driver sees a red stop sign and presses the brake. The event is ordinary, yet it contains most of the philosophy of mind. Light reflected from the sign changes the driver's sensory system. The visual process distinguishes red from other colors, an octagon from other shapes, and a sign from its background. Memory supplies a learned meaning. Together, these distinctions and that meaning become available to attention, planning, and action; motor signals depress the brake pedal; the car slows.

The same event has an inside. For the driver, the sign looks red, appears ahead, means stop, and becomes a reason to brake. The outside description follows photons, neurons, muscles, and motion. The inside description follows perception, meaning, decision, and action. This paper's core claim is that, for a conscious process, these descriptions meet in a constitutive identity: the complete organization of the internally implemented distinctions the process possesses and can use *is* its phenomenal character. To experience red rather than blue is to occupy a correspondingly different knowledge state. Experience is not a further event added after the discrimination becomes available to the process.

The argument follows a chain:

$$\text{difference} \rightarrow \text{process state} \rightarrow \underbrace{\text{centered knowledge}}_{\substack{\text{phenomenal character} \\ \text{within a phenomenal center}}} \rightarrow \text{action}.$$

An input selects among the process's possible states. The selected state carries information about that input. Information becomes knowledge for a process when the corresponding distinction is available within the process and can help govern its continuation. Within a phenomenal center, the brace marks an identity rather than a further transition. The complete organization of these distinctions—their similarities, contrasts, access relations, combinations, and causal uses—determines qualitative structure. No inner witness stands between knowledge and its effects.

A **center** is the functional locus at which a set of distinctions is jointly available and able to affect a common continuation. The definition is neutral with respect to phenomenality, and a center need not coincide with a person, an anatomical region, or a represented self. Its boundary follows causal-access relations at a chosen substitution level: information belongs to a center when a causal path makes its distinctions available there for use. A center that instantiates experience is a **phenomenal center**; its high-level state at a stage is an **observer-moment**. The access relations that constitute a center individuate the subject, determine which contents bind into one perspective, and explain why information present elsewhere in the same skull may remain outside it.

One thesis, its consequences, and its natural limit organize the paper. The **centered-knowledge thesis** is the core: if one center with a given complete centered-knowledge organization is phenomenal, then every center with an equivalent organization is phenomenal with the same phenomenal character, and that character is identical to the organization they share. From it follow phenomenal-functional covariance,

organizational invariance, and a criterion for consciousness-preserving replacement. The **phenomenal-minimum thesis** adds that some minimally sufficient organization, $K_{\min}$, marks the phenomenal boundary, without saying what it requires. The **minimum-boundary hypothesis** is what the identity yields when nothing is added to it: the first distinction a process can possess and use already suffices. It is the account’s null hypothesis about $K_{\min}$, and any richer threshold is a further postulate whose contribution must be specified. It is nevertheless stated as a separable hypothesis, in Section 7, because the core thesis can be evaluated in subjects whose consciousness is not in doubt without settling the boundary, and because the two are supported by arguments of different kinds. Readers who reject the hypothesis lose the paper’s verdicts about where consciousness begins, but nothing else.

1.2 Lineage and contribution

The thesis belongs to computational functionalism: mental states depend on implemented causal-computational organization rather than on one uniquely necessary substrate (Putnam, 1967; Piccinini, 2009; Chalmers, 2011). It has four close ancestries, each of which supplies a part.

The first connects consciousness with information poised for use. Dretske explains representation through the flow of information and later identifies experience with representational properties of perceptual states; Tye identifies phenomenal character with content poised for cognitive use; Chalmers proposes direct availability for global control as the cognitive basis of experience; and Kirk comes closest, emphasizing information that is *for* a system because it is immediately available in its discriminatory and decisional organization (Dretske, 1981; 1995; Tye, 1995; Chalmers, 1997; Kirk, 1994). The second explains qualitative character through relational structure: Clark’s quality spaces, Rosenthal’s discriminative roles, Palmer’s isomorphism constraint, and recent informational treatments of richness and ineffability (Clark, 1993; Rosenthal, 2010; Palmer, 1999; Ji et al., 2024). The third concerns centered knowledge: Lewis and Perry show that an uncentered description of the world can omit the self-locating knowledge that explains action, and Conee’s acquaintance reply to Mary distinguishes possessing every description of an experience from undergoing it (Lewis, 1979; Perry, 1979; Conee, 1994). The fourth is mechanistic work on the meta-problem: Metzinger’s transparent self-model, Graziano’s attention schema, and Frankish’s illusionism explain how a process can represent itself as aware and generate apparently ineffable judgments, and Cohen and Dennett argue more generally that empirically meaningful consciousness cannot be separated from function (Metzinger, 2003; Graziano, 2013; Frankish, 2016; Cohen & Dennett, 2011). Chalmers’s own organizational invariance, structural coherence, and double-aspect view of information, together with Zuboff’s replacement arguments and Marchal’s substitution-level framing, are the most direct precedents (Chalmers, 1995a;b; 1996a; Zuboff, 1994; Marchal, 2013).

The present contribution is not the discovery that consciousness has something to do with usable information. It is the proposal that these lines converge on one relation, and the working out of what that relation entails: that phenomenal character is the complete organization of a center’s usable knowledge; that the possessor of an experience is a causal center individuated by access rather than anatomy; that phenomenal judgment

makes epistemic contact with experience through causal provenance; that qualitative and discriminative structure cohere because they are one structure; that experience causes behavior without competing with physics, because the centered state is the difference-making variable at the level where the generalization holds; that invariance and replacement conditions are derived rather than assumed, and yield a criterion for which variables an emulation must carry, from which it follows that a connectome is not among the sufficient ones; and that the boundary individuating a subject is also its epistemic horizon, which is why physics presents itself to every observer as structure and dynamics and why the intuition of a hidden intrinsic nature beneath the relations arises. An event-level formalization of causal access supplies the tools, and the same formalization yields contrasting predictions against global-workspace, higher-order, and integrated-information theories.

Metaphysically, the thesis is a realist identity theory, and its relation to Chalmers's taxonomy of materialisms is best stated by saying what the taxonomy sorts by (Chalmers, 2003a). It sorts views by how they accommodate the explanatory gap: Type A denies that there is one, Type B accepts it and denies that it is ontological, Type C promises that it will close. The present view does not accommodate the gap but explains it. Three questions that the taxonomy runs together must be separated. Does complete causal-functional organization fix phenomenal character? Yes, without remainder: that is the identity, defended abductively in the manner Block and Stalnaker defend scientific identities rather than by analysis of concepts (Block & Stalnaker, 1999). Are the phenomenal truths deducible from the physical description? As deducible as any truth stated by way of a theoretical identity, so nothing about consciousness is special on this axis. Can a reasoner come to occupy a phenomenal state by reasoning from a description? No, and for a reason that follows from the present framework. A distinction selects one alternative to the exclusion of all the others, and an observer-moment is a complete organization of such distinctions, so to occupy one complete state is to occupy no other (Section 3.5). A complete description of an experience therefore cannot, for anyone, be that experience. This is the gap, and it is not a gap in the truths but in what any one center can occupy, and, for the center that occupies a state while holding its description, in what the state reveals of its own realization, which is nothing (Section 3.3). It follows that the identity's apparent contingency is derived, not conceded as a brute necessity in the way Chalmers takes Type B to concede it. The zombie is conceivable only in the sense that a bacterium's being merely organized, and not alive, was conceivable to a vitalist. Once the theory is understood the scenario is incoherent (Section 4.3). Phenomenal-concept accounts, and especially Balog's proposal that a phenomenal concept contains a token of the state it refers to, describe the conceptual shadow of this asymmetry, Hill and McLaughlin's appeal to distinct cognitive faculties its psychological shadow; Conee's acquaintance reply to Mary describes its epistemic shadow (Loar, 1990; Papineau, 2002; Hill & McLaughlin, 1999; Carruthers & Veillet, 2007; Balog, 2012; Conee, 1994). The present account gives the asymmetry itself a causal-computational form as occupation. Levine's demand, that an identity make intelligible why this organization has this character, it meets in part and declines in part (Levine, 1983). The structural features of a quality's character, where red stands among the other colors and what it does, are intelligible

from the organization (Section 6); the demand for a further, non-structural why is declined, because on the identity there is no further fact; and the persistence of the demand is explained rather than credited, by opacity and exclusivity, which keep the identity from being recognized from within even by a subject who occupies the state and knows its organization (Section 9.4).

The paper proceeds as follows. Section 2 develops the framework of processes, substitution levels, access graphs, and centers. Section 3 states the centered-knowledge thesis through the meta-problem, information for a process, immediate knowledge, and causal provenance. Section 4 derives invariance and replacement and treats zombies. Section 5 applies center individuation to binding, divided brains, and nested centers. Section 6 assembles the abductive case and confronts the alternative on which centered knowledge is merely the functional correlate of experience. Section 7 states the boundary hypotheses. Section 8 applies the account to Mary, Nagel, Searle, privacy, and other minds; Section 9 to what a center can know about its own history, program, and world. Section 10 compares the account with rivals and proposes tests. An appendix consolidates terms and falsification criteria.

2 The computational observer

Before asking what an observer can know, we need the right unit of description. A mind is a process, not a frozen arrangement of matter.

2.1 Processes, levels, and substitution

At any selected level, a computer changes from one state to another. Let x_k be its state at stage k , u_k an input available at that stage, and δ the rule determining what follows:

$$x_{k+1} = \delta(x_k, u_k). \quad (1)$$

The index k counts stages in the chosen description, and these need not coincide with clock cycles, neuronal spikes, or physical instants. A computer multiplying two thousand-digit numbers is, at the level of a mathematical library, performing one operation, while the virtual machine executes many bytecodes and the circuitry passes through vastly more electrical changes. A high-level state can therefore persist while lower-level states change, and it is identified by its causal role: what initiated it, what information it preserves, which continuations remain possible, and what it produces. A neural observer-state likewise extends across many milliseconds and brain regions, and only a fraction of its contributing neurons need fire at any instant, just as changing notes realize one melody. Consciousness therefore needs no permanently active material core.

Implementation is a relation between an abstract transition structure and a lower-level causal history in which earlier states and inputs produce later states in the required way, including under relevant alternative inputs (Piccinini, 2015; Chalmers, 1996b). An **interface** specifies the operations and exchanges available to a higher-level process: inputs, outputs, memory relations, timing, and possible transitions. A **substitution level** is an interface beneath which implementation may vary while these relations

remain fixed. The Java virtual machine is such an interface for a Java program: the program encounters the virtual machine and never the hardware beneath it, so the hardware may change without altering the Java-level computation unless the change alters an accessible result, exception, or timing signal. Marchal uses “substitution level” for the level at which functionally correct digital replacement preserves experience, and Chalmers’s principle of organizational invariance makes the parallel claim that systems with the same sufficiently fine-grained functional organization have qualitatively identical experiences (Marchal, 2013; Chalmers, 1995a).

The **substitution principle** states the preservation condition: details below an adequate substitution level may vary, provided every distinction available for use by the process, and every transition those distinctions can govern, remain fixed. This is the grain at which multiple realizability, in Putnam’s sense, holds (Putnam, 1967). The interface thereby acts as an **epistemic firewall**: if one implementation can replace another without changing anything the process can acquire, compare, remember, or use, the difference lies outside its point of view. While lower-level differences still implement and causally support the process, they do not become contents of its perspective. Section 9.4 draws the general consequence: what realizes a center’s distinctions is not merely unknown to it but unknowable from any position within it. The firewall and the stopping rule of Section 2.3 do different work: the firewall says that whatever lies beneath an adequate interface is outside the center’s point of view, and the stopping rule says where that interface lies, by difference-making under intervention rather than by what the center can introspect.

This causal criterion steers between two errors that pull in opposite directions. Identifying a mental state with one instantaneous microstate is too strict, since it rules out genuine multiple realization. Mapping an arbitrary physical history onto a formal state table is too permissive, as Putnam’s critique of machine-state functionalism showed (Putnam, 1988). Causal and counterfactual organization distinguishes an implementation from a retrospective labeling (Chalmers, 1996b; Zuboff, 1994). Ma and Kanai sharpen the point: a computational organization must be invariant under structure-preserving relabelings and grounded in variables that constrain one another under intervention, and a consciousness-preserving replacement must preserve internal state individuation, transition structure, and intervention profiles rather than agreement at a gross input–output boundary (Ma & Kanai, 2026; Kanai & Ma, 2026).

2.2 The access graph

Let the directed graph

$$G = (V, E) \tag{2}$$

represent a process’s causal-access organization. V is a set of causally individuated, state-bearing processing events, described at whatever grain the analysis selects. The graph may be finer than any center within it requires, and Section 2.3 gives the rule that recovers each center’s adequate substitution level from it. An event need not be instantaneous; it may occupy an extended spacetime region and be realized by many lower-level changes. When the same physical component participates at different stages, each participation is a distinct vertex. An edge $v_i \rightarrow v_j$ belongs to E when changing the state realized at v_i , under an appropriate intervention, could directly change the

state or transition realized at v_j (Woodward, 2003). A directed path marks a route by which a distinction realized in one event can become available to later events. Without such a path, a difference cannot affect the process’s state or continuation.

Assuming no closed causal curves, the event-level graph is acyclic. This is the standard structure in both of the fields from which the framework borrows. In distributed computing, the events of a computation form a strict partial order under Lamport’s happened-before relation, a relation of potential causal influence in which the successive events of one process are totally ordered while the whole remains acyclic (Lamport, 1978). In physics, causal set theory takes spacetime itself to be a partially ordered set of events under causal precedence, and because simultaneity is frame-relative a partial order rather than a global snapshot is the physically correct description of a process extended in spacetime (Bombelli et al., 1987). Acyclicity does not eliminate feedback from the implemented architecture. Grouping event vertices by the persisting components that realize them produces a coarser **component-level projection**, which may contain cycles or be strongly connected. A feedback path $A_t \rightarrow B_{t+1} \rightarrow A_{t+2}$ reaches a later state of A , never the event A_t from which it began. Its apparent cycle arises only when temporally distinct events are grouped by the component that carries them. A feed-forward unfolding can therefore preserve an equivalent labeled event graph while assigning successive events to different components (Doerig et al., 2019; Kanai & Ma, 2026). Figure 2 in Section 5.1 shows two cyclic component-level projections and the acyclic event graph they realize. Throughout, “recurrence” and “feedback” refer to properties of the component-level realization unless explicitly qualified.

Nothing fixes the outer scale of G . It may represent a selected process or, at the limit, the causal organization of the entire world. Expanding the represented domain adds events and relations but does not alter the organization that individuates a center already within it.

2.3 Centers and centered knowledge

Within G , a center C is a causal event-subgraph in which a set of implemented distinctions is jointly available and able to affect a common continuation. At a given stage, its **centered knowledge** K_C is the complete organization of the distinctions carried by that subgraph, their access relations, and the transitions they can govern. Centered knowledge is factive with respect to state-possession: C actually occupies the state whose organization is K_C , whether or not that state is veridical. No reflective self-attribution is required. Information may be encoded elsewhere in the system while remaining absent from K_C : a dormant memory, like a closed book, lies outside a center’s current state until retrieval brings it in, whereas a memory that already shapes recognition, expectation, or emotion has already crossed the boundary. These definitions characterize causal and informational organization without presupposing phenomenality.

Two clarifications fix what K_C contains. First, counterfactual relations enter K_C only when borne by the organization realizing the center’s present state, and only over the horizon that the substitution interface defines. A discriminator currently registering red still carries the distinction from green, because had green been present its state, and therefore its next transition, would have differed. What counts is the

set of transitions the present state can govern through the interface. Open-ended futures that would require further inputs to arrive are not part of the present K_C . This matters later (Section 10.4): a recurrent circuit and its finite feed-forward unfolding have equivalent K_C at each stage because the present state governs the same next transitions through the same interface. What the one purchases with temporal depth the other purchases with circuit depth, and neither continues without paying. Second, Maudlin’s Olympia shows why idle machinery must be excluded: a backup that would take over on another input supplies no present distinction merely by being attached (Maudlin, 1989). The rule is that a counterfactual dependency belongs to K_C when it is carried by the transition function of an event that is currently occurring in C . An ordinary conditional satisfies it: the comparison now being made would have sent control the other way, so the dependence on the condition is carried by a current event even though the branch not taken contributes none. Machinery whose engagement requires a different input to have occurred earlier realizes no current event in C and contributes nothing to the current K_C , although if engaged it may alter a later one. This shares Kanai and Ma’s exclusion of idle backups while defining participation through events rather than persistent components (Kanai & Ma, 2026).

The center is fixed by implemented dynamics rather than by an analyst’s chosen boundary. The access graph and an adequate substitution level jointly supply the **center-individuation criterion**: the graph identifies the event paths through which distinctions become jointly available and enter a common continuation, and the substitution level separates that organization from lower-level differences that can vary without altering it. The two notions are defined together, so a stopping rule is needed to prevent circularity. Fix **anchor events**, the events at which a continuation of interest is selected. The center is fixed by the implemented dynamics; the choice of anchors is a choice of which center to focus on, before the method traces its boundary. The adequate substitution level for the center containing them is the coarsest interface that preserves every intervention-sensitive dependency reachable backward from those anchors, and the center is the event-subgraph so reached. Coarsening further would omit a dependency, and refining further would add nothing available to the anchors. Two features of the rule keep it from running back to the beginning of the universe. Tracing adds a variable further back only when it preserves a dependency of the anchors that the nearer variable loses: in the center of Figure 1, tracing from the action selector reaches the joint register and the two detector outputs and stops there, whatever lies behind the detectors, because each detector output already carries every difference among pixel patterns that the selector can respond to. And it counts only variables whose alternatives select among the process’s continuations, so an enabling condition such as a power supply, whose alternatives run or halt the process rather than steer it, belongs to the implementation and not to the center. The two boundaries are the ones Dennett drew in assigning content to a neural state by its afferent and efferent connections together, what produces it and what it enables (Dennett, 1969). His intuition that awareness lies somewhere between a process’s inputs and its outputs is here given a location: tracing begins at the anchors, where continuations are selected, and stops where values enter, so the center is what lies between what a process receives and what it does, and the identity places experience in the organization of

that interval at the adequate interface. If several interfaces are adequate, they preserve the same dependencies by definition and therefore fix the same K_C , so nothing in what follows depends on the coarsest one being unique. Whether a given empirical model has located this fixed point is an open question for any real brain, and the account does not pretend otherwise. What it supplies is the condition a model must satisfy.

To count centers within a represented domain, first separate causally disconnected regions. Within each region, identify candidate continuations and trace their difference-making causes backward to the represented boundary or the adequate interface. A candidate qualifies only if it contains at least one implemented distinction whose alternatives can alter a continuation available to that candidate; when several distinctions are attributed to one center, they must be jointly usable in a common continuation. Proximity, common containment, and events that make no difference under any relevant intervention do not enlarge a center. Candidates need not partition G : a nested or overlapping subgraph counts separately when it independently satisfies the criterion, and their union counts additionally only when relations between them make their distinctions jointly available to a common continuation. Section 5.3 returns to what this permissiveness costs.

Several established insights fall into place. Minsky’s A-brain/B-brain makes the relation vivid: the A-brain connects to the world, the B-brain only to A, so “the A-brain is the B-brain’s world” (Minsky, 1986, p. 59). Two processes on one computer can share processor and memory while remaining isolated in protected address spaces (Dennis & Van Horn, 1966), while two components on different continents can implement one process. A biological organism may contain several centers at once. One center may also extend across anatomical or technological boundaries. These claims concern the individuation of centers and their states rather than persons, and so the account is compatible with different theories of personal identity and does not decide among them (Zuboff, 1990; 2025).

3 Epistemic contact and the centered-knowledge thesis

We can now follow the first three links of the chain: a difference selects a process state; the state becomes knowledge when the process can use it; and, for a phenomenal center, the process’s possession of that knowledge constitutes phenomenal character.

3.1 Phenomenal judgments as causal outputs

Chalmers distinguishes the hard problem—why information processing gives rise to experience—from the meta-problem: why we judge and report that consciousness poses a special explanatory problem (Chalmers, 2018, pp. 6–12). A wider phenomenal-judgment problem includes ordinary capacities: saying that one is in pain, distinguishing red from green, comparing two experiences, remembering how one felt, behaving as one does because of what one experiences. An adequate theory should explain not only why experience exists but how a process comes to know and act upon its phenomenal differences.

These capacities impose a causal constraint. Holding transition organization and inputs fixed, two cases identical in every causally available state cannot systematically

produce different continuations; in a deterministic process they produce the same continuation, in a stochastic one the same conditional distribution. A process can report red rather than blue, judge one pain stronger than another, or withdraw because of pain only if some corresponding distinction reaches a state capable of altering those continuations. The distinction need not be verbal or introspective, but it must be implemented within the process, available at the locus from which the outputs are governed, and causally usable there.

Return to the stop sign and replace the driver with an autonomous vehicle. Call the software that perceives, plans, and acts for it *the controller*; everything said of it below is said of an idealized controller that does whatever a maximally safe one must, and using it to illustrate a structural claim presupposes nothing about whether it is conscious. A prerecorded braking schedule may happen to stop one vehicle at one sign, but a controller that must stop for unexpected pedestrians, wherever they step out, must be sensitive to where each one is: different positions must put it into different internal states, and swapping its PEDESTRIAN-LEFT state for its PEDESTRIAN-RIGHT state must change what it does. Reliable competence in an unpredictable environment requires exactly this responsiveness to the variables that matter, and a controller without it is a more dangerous one. “I see red” and “I see blue” are likewise different outputs, and a causal explanation of why a process produces one rather than the other must identify a corresponding difference that reaches the judgment-producing machinery.

The constraint does not by itself establish an identity between experience and organization. A theorist might posit an experience that merely correlates with the judgment-producing state, or an additional phenomenal property that causes it. Either way the theory owes an account of *epistemic contact*: why the process’s phenomenal judgments track the experience they purport to describe. If the phenomenal difference makes no difference to the judgment-producing organization, the same process would make the same judgment without it and the judgment’s accuracy is unexplained. If it does make a difference, that difference must enter the causal organization as a usable distinction. Chalmers suggests that a first-order property grounding higher-order representation could play a key role in solving the meta-problem and might be a candidate for the physical basis of consciousness, leaving its nature open (Chalmers, 2018, pp. 39–40). The centered-knowledge thesis offers a specific candidate.

3.2 Information for a process

Shannon’s theory measures statistical information once a source alphabet, a distribution, and channel variables are specified (Shannon, 1948). It does not say what a signal means, which physical differences count as symbols, or at which level those differences matter. The string 01000001 denotes 65 as a binary number, the letter A under ASCII, and indefinitely many other things under other segmentations, codes, and decoding functions. Nothing in the string selects one of them.

An interpretation becomes operative when a process responds differently to the relevant alternatives. An ASCII parser, an image decoder, and a motor controller can receive the same bit pattern while using its distinctions differently, because each process’s transition organization fixes what the pattern does there. Information in this operative sense is always information *for* some process at some level, and the relativity

does not make content arbitrary: once process and level are fixed, counterfactual transition structure objectively constrains which differences the process can detect and what they can do. Bateson called information “a difference which makes a difference” (Bateson, 1972, p. 453). Here the second difference is indexed to the process whose future the first difference helps select. A scratch on a tape or a ring in a tree may let an observer infer its cause, but it is information for a process only when some process takes the distinction up and branches on it; until then it is potential evidence for any interpreter and operative content for none.

Human knowledge includes language, justification, and concepts. **Minimal functional knowledge** is the relation beneath them. A process minimally knows a condition when relevant alternatives reliably select different implemented states at the level in question; the selected distinction persists or propagates long enough to enter the process’s organization; and the process uses that state to select among its possible continuations. Reliable does not mean infallible: a bag blowing across the road from the left may place the controller in PEDESTRIAN-LEFT, and the state is then definite for the controller while misrepresenting its source, as Aristotle observed that one may be mistaken about what a white object is while the perception of white remains definite (Aristotle, 1984, *De Anima* III.3, 428b18–25). The controller’s states are information for that controller not because an analyst labels them but because its organization selects among them and uses their differences to steer.

3.3 Immediate knowledge and epistemic opacity

Chalmers uses **immediate knowledge** for the way a system may treat its presence in certain states as known without inference from further knowledge: asked why it says a green object is present it may cite its perceptual representation; asked why it is in that representation it may have no further reason to give (Chalmers, 2018, pp. 23–25). Two relations must be distinguished. First-order possession consists in being in a state that distinguishes green from its alternatives and makes that distinction available for use. Higher-order knowledge consists in a further state representing that the first is present. Chalmers’s discussion primarily concerns the second, whereas the centered-knowledge account begins with the first. A center need not represent “I am in the green state” before the green-selective state can guide recognition, comparison, memory, or action. A reflective state can make possession explicit, but it does not create the first-order distinction retroactively.

Possession of a distinction also differs from knowledge of its implementation. A visual center may know red rather than green without knowing which neurons fired or how the distinction is encoded. Its knowledge can be exact at the substitution level at which red and green guide different continuations while remaining radically underdetermined below. The controller exhibits the same relation between its parts. The module that plans its trajectory, *the planner*, receives “light is red” as a settled value from the module that classifies the camera image, with no access to the pixels, weights, or activations that settled it. Fodor’s informational encapsulation is this relation between modules (Fodor, 1983), and the stopping rule of Section 2.3 is encapsulation seen from the planner’s side: tracing back from its continuations stops where the value enters. Chalmers illustrates this with a system that can locate its state within an information space while lacking

access to its neural realization; asked what separates red from green, it can say only that they are “just different” (Chalmers, 1996a, pp. 290–292). That answer expresses knowledge of the contrast together with ignorance of its implementation. What matters within a phenomenal center is therefore whether a distinction is currently implemented in K_C and able to affect its continuation, and beliefs are no exception. A standing belief that makes no difference to the center’s current organization lies outside K_C , but a belief that is shaping recognition, expectation, inference, or emotion has entered the operative state and is part of what the center presently knows.

3.4 From first-order judgment to third-order mystery

Chalmers distinguishes three orders of phenomenal judgment: first-order judgments about the object of an experience (“That’s red!”), second-order judgments about a particular experience (“I’m having a red sensation now”), and third-order judgments about experience as a type (“Sensations are mysterious”) (Chalmers, 1996a, pp. 174–176). The orders classify judgments, not degrees of phenomenal richness. A first-order judgment is also distinct from the discriminative state that grounds it: the state distinguishes red from green and guides the process, while the judgment is a further state about the object discriminated.

A constructive route from the first to the third can be sketched, and Section 10.4 turns it into a test. The aim is a system that, in the course of its ordinary functioning, comes to make novel, internally generated judgments about its own experience, and the autonomous vehicle supplies the route, with each capacity arriving as the remedy for the accidents a controller without it would have. Begin with a controller that only discriminates and acts: each detected condition selects a response, a pedestrian detection brakes, a red octagon stops. Every judgment it makes is first-order, a judgment about the road. It brakes as hard for the bag as for the child, and with the same conviction when its classifier is certain as when it is guessing, because nothing in it registers the difference. The remedy is a confidence level attached to each classification. That is a state about the controller’s own current state, and it is already reflective in a modest sense, since what it tracks is the reliability of the machinery that produced the classification, the controller’s own fallibility. With it the controller distinguishes not only a pedestrian from a bag but being in the pedestrian state from not being in it, and can slow, look again, or override on that basis. Here the second order begins: a nonlinguistic monitor that treats its own present state differently from an absent or remembered one already implements a higher-order distinction, and its outputs are judgments about a particular current state rather than about the road. Training across many episodes then gives the monitor its standards, and the controller comes to register that a current state differs from a remembered one, that different situations can put it into the same state, as the bag and the child both did, and that some of its states are directly available to it while others are known only by inference.

A controller with all of this still has accidents, and they are of a distinctive kind. If it treats other vehicles as it treats rolling balls, as objects with trajectories, it will be surprised by them, because they respond to what they see, and what they see includes the autonomous vehicle. It brakes early at a junction to be safe and thereby invites the waiting driver to pull out. The controller that minimizes collisions is the one that weighs

the other driver’s response to its own candidate actions (Sadigh et al., 2016), and to do that its model of the world must contain a model of itself as one of the responsive systems in it. This is the condition Dawkins conjectured consciousness requires, a simulation of the world complete enough to include a model of itself (Dawkins, 1976, ch. 4), and it is reached by the route Carruthers argues our knowledge of our own attitudes takes, mindreading turned upon oneself (Carruthers, 2009). The next accident is subtler: a controller that predicts another vehicle from the state of the world rather than from what that vehicle can see will be wrong exactly where it matters, since the vehicle whose view of the signal is blocked by a truck will not stop for it. Predicting agents requires tracking what each of them can access, which is the false-belief task (Wimmer & Perner, 1983) arising in traffic, and it gives each agent in the controller’s model an epistemic horizon. Turned on the vehicle itself, the same competence gives the controller a horizon of its own, since a safe controller must represent what it cannot see, the pedestrian who may be behind the parked van. The model has by now acquired, without anyone designing them in, most of the relations this paper is about. The other vehicles’ states are private to it, reachable only by inference from position, heading, and what their sensors can cover. Its own state is not inferred at all: it is occupied. The model is a map of centers and their horizons on which every entry but one is filled in from the third person, and that one is filled in from the first.

The third order arises when the controller turns the same competence on itself. Having learned that it drives more safely the better it models itself, it has every reason to improve the model, and an idealized controller can do what no animal can: retrieve its own source code, its weights, and its engineering documentation, and study them. It then finds everything a third-person account could contain, which classifier fires for a red octagon, how confidence is computed, why the bag was mistaken for the child. Yet something strange happens, or rather fails to happen. Reading the complete description of its red-selective state does not place it in that state, and nothing in the description tells it what it already knew before it read a line of code, which alternative it currently occupies. Epistemic opacity has been in place from the start: the controller knew that the sign was red without access to the process that settled it, and the documentation, however complete, does not close that gap but redescribes it. Occupying a state and holding a description of it remain two different knowledge states, however complete the description. A sufficiently general reasoner can represent this as a problem: why can I tell red from green immediately when I cannot say how I do it? Why does parsing a complete trace of my red-selective state not place me in that state? Why is my own state available to me in a way the other vehicles’ states are not? Generalized, these are third-order judgments, judgments about experience as a type, and the first of them is a phenomenal judgment the controller can make but cannot explain. The architecture does not predetermine which philosophical conclusion the system will draw—ineffability, self-access limitation, or deflation—but it explains why experience becomes an object of inquiry and why the difference between occupying and describing a state can appear to reveal an explanatory gap. Nothing in the goal of driving safely aimed at judgments about minds, yet they can arise in a controller that must model the other drivers’ processes and, to drive among them, its own. This is an account of how the meta-problem arises. The thesis adds the constitutive claim: within a phenomenal center, the

relation that third-person description seems to omit is not a causally separate property. It is the process-relative relation in which an implemented distinction is possessed and guides the center from its own locus. Note that nothing here makes reflection or memory necessary for phenomenality. An infant, animal, or nonreflective process might possess rich first-order distinctions without wondering why consciousness exists.

3.5 Causal provenance, occupation, and description

An output can be accurate by accident. A recording may happen to play “I see red” when a red object is present, yet nothing in the mechanism producing the words discriminates red from its alternatives, so the words express no judgment at all. Consider instead the controller. Red and green signals place its perception in different states; those states enter a process that can compare them, retain their relations, guide action, and, on the display that narrates the vehicle’s decisions to its passengers, report them. When the display reads “stopping for red light”, the causal explanation passes through the red-selective state; substituting the green-selective state changes what is displayed and what is done, and preventing either state from reaching the planner removes the competence. This counterfactual dependence is the report’s **causal provenance**. Provenance need not begin with a current external stimulus: a remembered color can ground a comparison, and a hallucinated red state can ground an accurate judgment about one’s current experience while misrepresenting its source. External accuracy and epistemic contact with the current state are separable. Three things are therefore distinct, and the vocabulary should keep them so: occupying a state, which is actual whether or not the state is accurate; the state’s representing its external cause correctly; and a judgment about the state being correct about it. Provenance and occupation secure the first and give a judgment access to its target; whether the judgment gets the target right is the competence of the monitoring organization that produces it, which is part of K_C and fallible. Speech is one downstream use; epistemic contact may be expressed through attention, comparison, memory, expectation, or action.

Given a phenomenal center C , *first-person occupation* is C ’s relation to the distinctions that make up K_C : they are present in the very organization from which C ’s continuation is governed, rather than described to it from elsewhere. The center need not form the thought “these are mine,” and there is no further observer inside the center inspecting them. A third-person observer can describe those distinctions completely—their implementation, computational role, access relations, and every continuation they can govern—but the observer’s representation of C is part of the observer’s own K_D , not of K_C . When C reports an experience to D , the report creates in D a state *about* C ’s experience, which may allow D to classify it, predict C ’s behavior, or construct a corresponding state from D ’s own resources; it does not place D in C ’s present state. If D is reconfigured to implement a corresponding organization, D may come to occupy a corresponding state, but because D now realizes it, not because a description conveyed a hidden object. If a direct interface makes a distinction jointly available to both processes, their access graphs overlap or form a wider center, and the case is no longer ordinary communication. The first-person/third-person distinction thus requires no two ontologically separate events: it identifies two positions within one causal organization.

Occupation has a consequence that follows from the definitions and will be used repeatedly. A distinction selects one alternative to the exclusion of all the others, and an observer-moment is a complete organization of such distinctions, so to occupy one observer-moment is to occupy no other, and no center can occupy them all. Call this the **exclusivity of occupation**. It is why a description that fixes every truth about an experience cannot, for any reader, be that experience: not because a truth is missing, but because occupation is not a truth. Section 8.2 develops its consequences, and the Mary, Nagel, and Searle cases of Section 8 apply the distinction between describing and occupying. None of them introduces a further kind of fact.

3.6 The centered-knowledge thesis

For any center C , K_C was defined without presupposing phenomenality. The thesis adds a constitutive claim that begins with any organization that possesses phenomenality and applies to every equivalent instance:

The centered-knowledge thesis. If one center C with complete centered-knowledge organization K_C is phenomenal, then every center C' with an equivalent organization is phenomenal and has the same phenomenal character, and that character is identical to the organization they share. Within each such center, the distinctions phenomenally present are exactly those belonging to its centered knowledge, and the phenomenal role of each distinction is fixed by its place within that organization.

Compactly,

$$\begin{aligned} P_C &\equiv K_C \quad \text{for phenomenal } C, \\ C \text{ is phenomenal} \wedge K_C \cong K_{C'} &\implies C' \text{ is phenomenal} \wedge P_C \cong P_{C'}, \end{aligned} \tag{3}$$

where $\equiv$ denotes constitutive identity and $\cong$ equivalence of complete organization at the relevant substitution level. P_C is C 's phenomenal character, the total structure of what it is like to be C at that stage, and K_C is that same organized state described in computational and epistemic terms. The first line, the identity, says that these are one thing under two descriptions, so phenomenal character is not merely fixed by the organization but is it. The second line, the equivalence clause, is its corollary: if the character is the organization and not something the organization produces, then a center with an equivalent organization has that character, and to have a phenomenal character is to be phenomenal. The restriction to phenomenal C in the first line says which organizations are phenomenal characters; it does not leave open whether an equivalent of one is. To deny the clause while granting the identity is to hold that phenomenality belongs to the instance and not to the organization, which is the correlate view that Section 6 confronts; Section 10.4 states the clause again, as P2, in a form that does not depend on the identity. The thesis fixes character by the state the center possesses, not by its external cause: a veridical perception, an illusion, and a hallucination share phenomenal character when they place the center in the same K_C .

Because the identity is between phenomenal character and the complete organization, any difference in one is a difference in the other. This is the covariance principle:

Phenomenal–functional covariance. Among phenomenal centers, phenomenal character varies exactly with complete centered-knowledge organization. A non-equivalent difference in that organization entails a phenomenal difference, and a phenomenal difference requires a non-equivalent difference in that organization.

Call any non-equivalent difference in complete centered organization a Δd ; the symbol denotes an organizational difference, not a scalar magnitude. Variation beneath the substitution level, or a relabeling that preserves the complete organization, is not a Δd . Within the phenomenal domain, models and interventions can then identify which organized relations constitute particular qualities, turning the thesis into a constructive program for explaining phenomenal richness: a human color experience occupies a relational space of similarities, contrasts, locations, constancies, and possible comparisons; pain combines location, intensity, attention, aversion, learning, and control—neither an isolated withdrawal reflex nor an actor’s scream establishes that full organization.

3.7 Valence

Pain exposes a dimension that sensory examples can obscure. **Valence** is the way an occupied state matters to the process that possesses it: a state is positively valenced when possessing it disposes the process toward a condition and negatively valenced when possessing it disposes the process against one. Representational content concerns what is present to a process, while valence concerns how possession of that content reorganizes its possible continuations. Minsky argued that hurting consists in the cognitive changes themselves, a “complex rearrangement of our disposition of resources” (Minsky, 1998). The thesis accepts this and draws its phenomenal implication: within a phenomenal center, the reorganization constitutes how the state feels. Pain is therefore not a magnitude stored in some variable. It is a change in how the whole process proceeds, in what it attends to, remembers, learns, prioritizes, and does, and differences in that *how* are the differences among feelings.

Negative valence is an organization in which possession of a state characteristically redirects the center toward stopping, escaping, reducing, or preventing a condition, while altering attention, memory, learning, and expectation accordingly, and positive valence reorganizes the center toward preserving, approaching, repeating, or recovering one. Pain, fear, hunger, grief, and frustration are all negatively valenced yet differ phenomenally because they occupy different places within the complete organization. What a valenced state changes is the center’s priorities and the continuations available to it, and competing states can override the action it would otherwise favor, so a person may endure pain for a goal, refuse a pleasure, or freeze under threat. For the same reason a reward variable, loss signal, or damage flag does not by itself establish pleasure or pain, but contributes to valence only through its place within a larger organization of priorities, learning, and control. Undergoing pain also differs from representing it: a doctor, actor, or model may accurately represent another process’s injury without

occupying the corresponding valenced organization. Conversely, a process need not possess a reflective concept of itself for a state to matter from its own locus.

3.8 The hard problem and the reality of experience

Chalmers formulates the hard problem as explaining why physical processes give rise to experience and why particular experiences have their character (Chalmers, 2020). The thesis answers by saying what experience is. The answer is a constitutive identity, defended abductively in Section 6 rather than deduced from premises every party accepts, and it runs as follows. A phenomenal center’s knowledge-bearing process is not first completed and then accompanied by an additional event. When an implemented distinction becomes available to that center and governs its activity, the center’s occupation of that state is the phenomenal fact. Described from outside, it is a causally effective discrimination. From the operative position of the center, it is knowledge possessed. These are two descriptions of one event, not two events awaiting a psychophysical bridge, and “accompanied” conceals the duality the theory rejects. If a phenomenal center’s processing is described without reference to the center—as signals, transitions, and behavior—experience appears missing from the description. Restoring the omitted relation reveals which distinctions the center possesses and how they enter its activity. The phenomenal inside is not a further object observed by an inner witness but the presence, for that center, of one alternative rather than its incompatible alternatives in the state that guides the process from that locus.

It is fair to ask what this realism adds beyond the label, since illusionists accept the same mechanisms and call the resulting properties quasi-phenomenal (Frankish, 2016), and attention schema theory supplies the most developed such mechanism, a simplified internal model of attention whose contents the system reports as subjective awareness (Graziano, 2013; Webb & Graziano, 2015). The difference is a claim about reference. Illusionism holds that introspection represents experience as having intrinsic, ineffable, non-functional features that nothing has, so that phenomenal concepts as ordinarily deployed are empty. The present view holds that phenomenal concepts are recognitional and demonstrative, deployed from within an occupied state and referring to the organization occupied; what introspection presents as immediate, private, and hard to describe is exactly what centered knowledge is—possessed without inference, unavailable through description, and specified by relations too numerous to verbalize—so that no error need be attributed to introspection itself; the one error the account attributes is the inference from the felt intrinsicness of occupation to a non-structural residue (Section 9.4). Frankish argues that such conservative realism must collapse into illusionism or into a substantive non-functional view. It collapses into neither as long as the introspective appearances are explained rather than explained away, and Sections 3.3, 3.4, and 3.5 were that explanation, completed by Sections 8.2 and 9.4. Where the illusionist says that there seems to be something it is like but there is not, the present view says that the seeming is the being: to occupy a centered state is for there to be something it is like, and that is not an appearance of anything further.

Rejecting the thesis incurs a specific burden: explain how a phenomenal center’s complete knowledge organization could remain fixed while its phenomenal character differed, or was absent, when the alleged difference is by hypothesis absent from every

comparison, judgment, memory, and action through which the center could know its own experience. Section 4.3 develops that burden.

4 Consequences of the identity

Three things follow once the identity is stated: invariance under replacement, a place for mental causation, and the incoherence of a functionally identical zombie. This section takes them in turn.

4.1 Organizational invariance and replacement

Because Equation 3 applies across equivalent complete organizations, another center that instantiates the same K_C is itself phenomenal with the same character. A difference helps specify P_C only if it enters or alters K_C ; differences sealed behind the epistemic firewall leave P_C unchanged. Organizational invariance therefore follows from the identity rather than entering as an independent psychophysical principle.

Chalmers formulates the preservation claim directly: given a conscious system, “any system that has the same functional organization at a fine enough grain will have qualitatively identical conscious experiences” (Chalmers, 1995a, sec. 1). His neural-replacement arguments support this as a nomological principle: if fine-grained organization is preserved, a substantial phenomenal alteration would be disconnected from the subject’s capacity to notice, judge, or report it. Zuboff’s hemispheric-replacement argument makes the same point: replace the machinery serving half the visual field while preserving its organization and its exchanges with the rest of the brain; if the replaced half then looks different in a way the person can compare or remember, the replacement did not preserve the organization. If the center’s entire epistemic situation is preserved, an alleged loss is detached from all the ways that center knows its experience (Zuboff, 1994). Zuboff names the view this argument targets *intrinsicism*: that properties of a bit of processing apart from its causal role, whether identified with the experience or held to generate it, are essential to its phenomenal character (Zuboff, 2008). In the present terms, intrinsicism is the claim that phenomenal character depends on variables beneath the adequate interface, the very variables the substitution principle allows to vary. Chalmers characterizes the relevant grain as fine enough to fix behavioral capacities but does not identify the substitution boundary for a particular process or say why preserving it preserves experience.

The thesis supplies both additions. The substitution principle of Section 2.1 turns the fine-enough clause into a criterion: an interface is adequate for C when preserving it preserves every constitutive distinction in K_C , every access relation through which those distinctions are available, and every transition, including relevant timing relations, that they can govern across the relevant counterfactual conditions. The required grain is the coarsest level at which this complete organization remains invariant, not a fixed anatomical scale. The criterion has practical consequences for brain emulation and uploading: a reconstruction need not reproduce every molecular detail, but it must capture or functionally reconstitute every variable whose variation could alter the center’s distinctions, access relations, or transitions. Preserving only some usable

distinctions might produce a conscious digital process, but a phenomenally faithful upload requires preservation of the complete K_C .

Pharmacology shows how demanding this is. Ethanol, ketamine, or an anesthetic leaves the connectome intact, changing nothing about which neurons connect to which, yet alters K_C profoundly by changing how those connections transmit: their gain, timing, and plasticity. Such variables therefore lie above the interface, since they alter transitions. Marder’s stomatogastric circuit made the general point: one fixed wiring diagram produces different rhythms under different neuromodulators, so the connectome is necessary but not sufficient (Marder, 2012; Bargmann & Marder, 2013). Sandberg and Bostrom’s roadmap for whole brain emulation asked at which of eleven levels of detail scale separation holds and left the question open (Sandberg & Bostrom, 2008). The identity answers it in kind rather than in scale. A point-neuron emulation with fixed parameters is too coarse to be adequate, because neuromodulation continually resets the very parameters it holds fixed, and the distinctions that resetting makes available to the biological K_C have nowhere to appear in the model. What an emulation must carry is not the molecules but the effective variables they realize, wherever among the scales those variables sit: the substitution level is a set of variables the transition function depends on, not an anatomical stratum. Which physical variables realize that organization remains an empirical task.

4.2 The fourth link: mental causation

The chain of Section 1.1 ends in action, and the identity must say how a centered state causes without competing with the physics that realizes it. Kim’s exclusion problem states the difficulty: if the neural event is a sufficient cause of the driver’s braking, a mental cause is either idle or an overdetermining second cause (Kim, 1998). The problem assumes two events, or, in its property-level form, two difference-makers. On the identity there is one event, and the next paragraph shows there is one difference-maker. The recognition of the stop sign and the visual and prefrontal activity realizing it are not successive events in a causal chain but one state described at two grains. The autonomous vehicle makes the structure inspectable, since every causal relation in it is known: the controller’s STOP state and the electronics that realize it are one state, and “the vehicle braked because it recognized the sign” and “the vehicle braked because these transistors switched” are one transition described at two organizational grains rather than rival causal chains. “Seeing the stop sign caused the driver to brake” is true in the same way and for the same reason. Causation unfolds across time, from recognition to decision to motor command, not across levels.

Which description is the causal one is settled by difference-making. In the interventionist sense, a variable is a cause when changing it under an appropriate intervention changes the effect (Woodward, 2003; 2015), and nothing confines causes to the lowest-level physical or chemical vocabulary. Changing the STOP state, in the driver or in the controller and however it is realized, changes the trajectory that follows. The higher-level variable is therefore the one that captures the invariant generalization: the organization is what the effect depends on, and the realizer is what happens to carry it. Sperry made the point for brains, observing that neurons do not decide independently when to fire but take their “firing orders for the day” from the organization

of the circuit that contains them (Sperry, 1965, p. 83), and his rolling wheel, whose organization constrains the trajectories of its molecules, supplies the physical analogy (Sperry, 1969). Hofstadter’s tangled hierarchies describe the same structure, in which lower-level activity realizes a higher-level state that then guides the process’s later lower-level activity (Hofstadter, 1979).

Two consequences follow for the argument to come. Experience is not epiphenomenal on this account: the phenomenal state is the knowledge-bearing control state, and it causes what it causes in exactly the sense the controller’s STOP state causes the braking. And epistemic contact needs no alignment a dualist must supply: the state that is experienced is the very state that produces the report, memory, and comparison that concern it, so a judgment reaches its object without a bridge, and its accuracy is the ordinary reliability of the organization that forms it (Section 3.5). A view on which experiences do not cause the judgments that describe them must explain the accuracy of those judgments some other way; Section 4.3 argues that none has.

4.3 Zombies and epistemic contact

A philosophical zombie, in its strongest form, duplicates a conscious person physically and functionally—every discrimination, memory, comparison, introspective judgment, report, and counterfactual disposition—while lacking experience (Kirk, 1974; Chalmers, 1996a). A merely behavioral imitation might be exposed by examining its causal provenance, but the zombie cannot be, since by hypothesis its provenance is identical. Let O denote the subject’s complete physical and causal organization and Q the phenomenal character the zombie allegedly lacks:

$$\text{conscious subject: } O \wedge Q, \quad \text{zombie: } O \wedge \neg Q. \quad (4)$$

Writing this is easy. Positively conceiving the two cases requires giving coherent content to the claim that Q is absent after every process by which the subject discriminates, attends, remembers, compares, and knows has been held fixed. Appending “and yet there is no experience” asserts that phenomenal character is separable from organization, but asserting a separation is not conceiving one.

Redescription does not eliminate.

When a person is described exclusively in the vocabulary of particles, neurons, or state transitions, it may appear that knowledge, red, and experience are missing because the words are missing. But a vocabulary’s omission is not an ontological subtraction. A finite instruction set can implement an unbounded range of programs whose properties are not named in the instruction set; describing the controller entirely as voltage changes does not eliminate the sign recognition it implements, and describing an organism entirely as chemistry does not eliminate the life it implements. Chalmers’s structure-and-dynamics argument holds that physical theory characterizes only structure and dynamics whereas phenomenal truths are not structural (Chalmers, 2003a). That final premise is what the thesis disputes: phenomenal character is the complete relational organization of the knowledge a center possesses—its contrasts, similarities, combinations, access relations, values, and possible effects—and need no more appear among the primitives of physics than sign recognition among the primitives of electronics. Section 9.4 adds that the

structural character of physical description is itself predicted by the account of points of view rather than being a limitation physics happens to have.

Epistemic symmetry.

Shoemaker's objection to absent qualia is that if qualia could be absent while all functional organization remained fixed, they would be irrelevant to everything a subject knows about her own experience, so that no one could know she had them at all (Shoemaker, 1975). Kirk develops the same problem (Kirk, 2008). The zombie duplicates every internally usable distinction and every operation on them. It sincerely judges that it is conscious, recalls experiences, attends to present qualities, wonders whether other beings have qualia, and develops the same philosophical arguments as its conscious twin, from the same evidence, with the same reasoning, confidence, and error checks. Dennett describes zombies as "wrong, in ways that neither they nor we could ever discover" (Dennett, 1995), and the problem runs deeper than external undetectability: the zombie cannot discover its error through its own reflection either, because it contains the same introspective mechanisms, memories, apparent demonstrations, and certainty. The symmetry cuts both ways. If the zombie's introspective states give it no contact with phenomenal properties, then the physically identical states of its conscious twin give her none either, since nothing causal distinguishes them, and we are no better placed to know that we are not zombies than the zombie is. Yet Chalmers must classify the subject's judgment as knowledge and the zombie's identical counterpart as false or unjustified. What breaks the symmetry?

The causally sealed annex.

Any reply that keeps the duplicate exact must answer that question while granting that Q makes no difference to anything the subject does, and every such reply therefore places phenomenality in a causally sealed annex. Inside are Q , the psychophysical laws that assign it, whatever acquaintance the subject is said to have with it, and whatever content or justification is said to distinguish genuine phenomenal belief; outside are every discrimination, memory, comparison, judgment, utterance, and experiment. The annex exposes a dilemma. On the causal-exclusion horn, Q affects attention, memory, recognition, comparison, judgment, or action; it then makes a causal difference within the process, the zombie no longer duplicates the subject's complete organization, and the effect belongs within the centered state the theory must preserve. On the epistemic horn, Q affects none of these, and then nothing the subject does can be explained by it. Interactionism opens the annex only by abandoning the exact duplicate. The zombie argument obtains its distinctive force only when Q is held apart from the complete causal organization, and that is also the condition under which Q loses its explanatory role.

Replies that concede causal idleness.

Two replies take the epistemic horn directly: they accept that Q does nothing and hold that the subject knows it anyway. On the first, the epiphenomenalist's common-cause reply, the experience and the belief are effects of one brain state and reliably co-occur under the actual psychophysical laws, and reliability is all that perceptual

knowledge requires too (Robinson, 1982). On the second, Chalmers's, the belief is knowledge because the quality is a constituent of it: the subject is acquainted with the quality by instantiating it, and the belief needs no causal sensitivity to be justified (Chalmers, 2003b); the zombie's physically identical state, having no quality to build its concept from, is empty or false. Put plainly, on either reply the conscious subject and the zombie do exactly the same things for exactly the same causal reasons, and the one is said to know and the other not because a causally idle ingredient is present in one and missing from the other. Neither reply escapes the horn, and it helps to see where each stops. Perception is reliable because the object causes the belief, so its reliability is explained by the relation between them, whereas co-occurrence under a brute law explains nothing about why the belief is correct. Constitution does better, but for one belief only: a direct phenomenal belief, "this quality," has the quality as a constituent and, Chalmers holds, cannot be false. A subject does far more with her experience than demonstrate it. She identifies red rather than green across occasions, judges that orange lies between red and yellow, notices that this shade is redder than the one a moment ago, corrects her misclassifications, and builds general theories of experience. These are capacities, achievements of a process that must be sensitive to what it is a capacity for, and they deploy what Chalmers calls standing phenomenal concepts, which persist when the quality is gone and which, on his own account, are not constituted by it: their content is fixed partly by dispositions to recognize its instances (Chalmers, 2003b). A disposition to recognize is a causal capacity. Under the closure the zombie argument assumes, it responds to the physical correlate of the quality and to nothing else, so the quality cannot shape when it fires or when it errs. Either the standing concepts track the quality, and closure fails, or they track the correlate, and everything the subject knows of her experience beyond the demonstrative moment is knowledge of the correlate. Constitution thus secures incorrigibility for one belief at the price of all the rest. The defender will grant that the zombie has every one of them, and that concession is the argument: everything the subject can do with her experience is then explained without Q , and a knowledge that contributes to none of it is a knowledge she cannot exercise. Agreement by stipulation explains none of it, not why the subject identifies red correctly, not why she corrects her errors, not why her theories of experience improve.

The Russellian escape.

One position partly escapes this dilemma, and it should be named. A Russellian monist holds that phenomenal or protophenomenal properties are the intrinsic natures of physical causal powers. On that view Q is neither causally idle nor an addition to the causal structure, so the causal-exclusion horn does not bite. But a Russellian who allows a *structural* duplicate with different or absent intrinsic natures faces the epistemic horn in full, since that duplicate makes every judgment the original makes, and one who identifies each quality with its power meets it at the interface instead, where carbon and silicon realizers of one organization differ in their powers and character must follow either the organization, which is the identity, or the realizers, which no judgment tracks. Sections 9.4 and 10.3 develop the comparison.

The retreat to psychophysical laws.

A naturalistic dualist, in Chalmers's phrase, treats phenomenal properties as fundamental and posits fundamental psychophysical laws assigning them to organizations, so that a zombie world shares our physical laws and history but lacks the psychophysical laws. This makes the modal description formally consistent once Q 's independence has been assumed, but it provides no evidence for that independence. If Q has no physical effects, radically different psychophysical laws predict exactly the same physical observations. Consider hypotheses on which all normally organized humans are conscious, on which only five people are, on which only people wearing blue are, or on which the reader of this sentence is not conscious now. Each predicts the same neural activity, memories, avowals, publications, and experimental results, so the physical evidence cannot select among them. Simplicity may make some mappings more attractive, but that is an additional commitment, and first-person evidence cannot break the tie, since every act of noticing, remembering, or using it is fixed by O and reproduced under every hypothesis on the list.

Chalmers's own replacement arguments exploit this difficulty. If qualia changed substantially while organization remained fixed, the subject could not notice, remember, or report the change; Chalmers treats this dissociation as strong reason to accept organizational invariance in the actual world (Chalmers, 1995a). These are his fading and dancing qualia, and the zombie is their limiting case: consciousness disappears completely while every judgment that it remains vivid continues unchanged. He restricts the conclusion to nomological necessity and allows that a duplicate could lack consciousness under different laws. The position is coherent, but the change of laws now performs all the work: the zombie exhibits no positively conceived deficiency of its own, since it is the same organization placed under a law that withholds Q . Unnoticeable phenomenal alteration is treated as implausible enough to support invariance in the actual world, yet total phenomenal absence with the same inability to notice is used to support metaphysical separability. The nomological/metaphysical distinction prevents a formal contradiction, but it does not turn the zombie into independent evidence against identity.

The ledger.

To keep the zombie exact, one must accept the following. A belief about one's own experience is about an ingredient to which no physical process has access, not even the process that forms the belief, and that ingredient fixes what the belief is about while making no difference to how the belief is formed, used, or reported. A duplicate that lacks the ingredient makes every judgment the original makes, and its identical beliefs are false or empty. Which subjects have the ingredient is settled by facts that no observation, and no act of introspection, could detect. And none of the subject's capacities to identify, compare, and correct her experiences owe anything to the ingredient. The three replies differ over what the ingredient is: a lawfully attached quality, a constituent of the belief, or the intrinsic nature of a causal power. They do not differ over these commitments. Each reply that keeps the duplicate requires every one of them; the Russellian who denies the duplicate is answered at the interface instead.

Result and scope.

Equation 4 shows only that no contradiction appears while O and Q are written as independent symbols. That is negative conceivability: nothing said so far rules the scenario out. Positive conceivability would require a coherent picture of what differs between the two cases, and none has been given, since every fact that could give the difference causal or epistemic significance is fixed by O , and Q contributes nothing beyond the instruction to call one case conscious. A functional duplicate occupies the same centered state in exactly the same sense as the original, and that is the reason it is conscious; to insist that it could occupy every such state while lacking what it is like is to reintroduce Q as a further constituent, which assumes the separability at issue. Chalmers's master argument holds that a physical account of our concepts of experience cannot also explain why the gap seems to exist (Chalmers, 2006). Its first horn, that an account a zombie would satisfy equally cannot explain our epistemic situation, assumes that our situation differs from the zombie's, which is what the ledger prices. The present view explains the gap once, for one case: a complete description of O never places its reader in the state described, so the reader can imagine the description satisfied and experience absent, when what is absent is only her own occupation (Section 3.5).

The zombie therefore provides no independent evidence that phenomenality is separable from complete causal organization; it merely denies the identity. Given the identity, $O \wedge \neg Q$ is a contradiction, since Q is then part of O , and the scenario is conceivable only as a bacterium's being merely organized, and not alive, was conceivable to a vitalist: conceivable but impossible in Chalmers's vocabulary, incoherent once the theory is understood in the present one. The argument works at two levels. Without assuming the identity, it shows that the scenario has no content bearing on anything a subject could know, so that its conceivability is not evidence. With the identity in hand, it shows the scenario incoherent, which supports the organizational-equivalence clause of the thesis. This is not a deductive proof against every possible metaphysics. It is a demonstration that the zombie intuition cannot be cashed as an argument without presupposing its conclusion, and the ledger states what that presupposition costs.

5 One brain, many centers

The access graph explains why the neural correlates of consciousness can be distributed and changing, why features bind without a theater, and why information inside one skull can remain outside the perspective called "I." Research on neural correlates seeks mechanisms sufficient for conscious contents (Koch et al., 2016), and the apparent unity of experience can mislead us into seeking a simultaneous snapshot in a localized group. A computational observer-state is instead an organized process whose events may be distributed across regions and time. Motion makes the temporal spread unavoidable: to distinguish movement a visual process must relate an object's position at one time to its position at another, so even a percept delivered as a single velocity vector is produced by a temporally extended process.

5.1 Binding is structured joint use

Suppose a red tail light glows beside an octagonal billboard, and the controller must stop for red octagons only. Separate detection of red and of the octagon does not establish which features belong together, and binding requires preserving the assignments (Treisman & Gelade, 1980). At the simplest level, two separately computed distinctions bind when a later process uses their conjunction with a common index: for an indexed object o , a stop-sign discriminator computes $\text{Red}(o) \wedge \text{Octagon}(o)$, so the scene must not be classified as containing a stop sign merely because red and an octagon are each detected somewhere in it. A small part of the controller makes the joint organization explicit. Let $r = 1$ denote red, $s = 1$ octagonal. A joint register $q = (r, s)$ selects one of four conditional transitions (Table 1, Figure 1).

Table 1 Joint states and conditional actions in a two-feature center

Joint state q	Registered alternative	Selected action
(0, 0)	Neither feature (a green circle)	Proceed
(0, 1)	Octagon, not red (a faded sign)	Slow
(1, 0)	Red, not an octagon (a tail light)	Keep distance
(1, 1)	Red octagon (a stop sign)	Stop

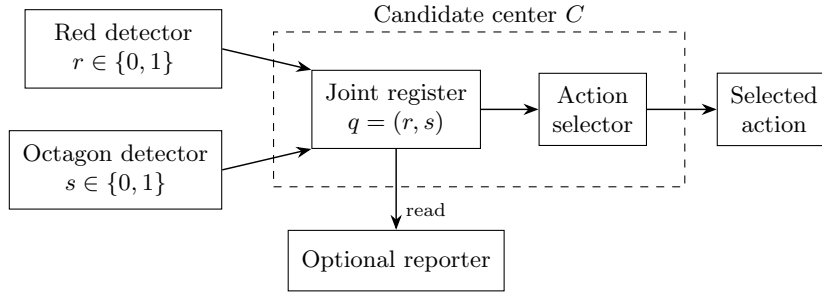


Fig. 1 A joint red–octagon center within the controller. The register and action selector form candidate center C ; both features enter its conditional continuation. An optional reporter reads the register without feeding back into C . Arrows represent causal influence; the dashed boundary marks the candidate center.

Changing either feature while holding the other fixed changes the selected transition. This center is richer than two isolated detectors: one process possesses four joint alternatives and uses their conjunctions to govern a common continuation. This is the **composability principle**: separately computed distinctions bind when a later process uses them together in one indexed joint state. Attention, spatial maps, object files, synchrony, or recurrence may implement such use; what matters, as Dennett’s Multiple Drafts model also holds, is how established content affects later activity—it need never be re-presented to a master discriminator (Dennett, 1991, p. 113).

Two relations should be kept apart. *Co-consciousness* is joint availability to one center: Aristotle observed that discriminating white from sweet requires both to be present to one and the same perceiver (Aristotle, 1984, *De Anima* III.2, 426b17–23), and within a phenomenal center joint availability is enough for two contents to belong to the same experience. *Feature binding* is the further assignment of several features to one object, which requires the common index. Both are relations within K_C , and both are required for coherent action: one process may register hunger and recommend grazing while another detects a predator and recommends flight, and only a common control state containing both distinctions, their urgency, and their incompatible recommendations can settle the conflict. Unity of consciousness is therefore closely related to unity of control.

Modules can achieve broad access either by exchanging outputs directly or through an integrating hub (Figure 2a,b): the autonomous vehicle’s camera, lidar, and radar modules may pass their detections to one another or to a fusion stage that each of them reads. The two architectures differ physically, but in each the same distinctions reach the same later events to guide the same transitions. Neither needs a homunculus. Dennett writes that “the brain is Headquarters” while denying any “deeper headquarters” (Dennett, 1991, p. 106). The access-graph account adds the positive claim that, through these relations, the brain-wide process itself possesses the distinctions and thereby constitutes a center. This is the **virtual-center principle**: a center can be distributed rather than confined to a privileged place. Both architectures contain cycles, but the cycles belong to the component-level projection. Unfolded into events (Figure 2c), each module’s participation at each stage is a distinct vertex and every edge runs forward in time, while the hub’s events appear only as relay vertices that add path length without adding distinctions. At the cognitive substitution level those relay events fall beneath the interface, and the two architectures realize equivalent event-level organization.

5.2 Divided brains and unconscious processing

Blindsight patients navigate obstacles and even hit or catch thrown objects while denying that they see anything (Weiskrantz, 1986; de Gelder et al., 2008; Lane et al., 2024). Split-brain patients act on information unavailable to the speaking hemisphere (Gazzaniga, 2000; Nagel, 1971). The apparent paradox dissolves when access is indexed to a process: visual information may guide one center while remaining unavailable to speech, episodic memory, or self-attribution in another. After callosal division the typically nonspeaking right hemisphere can understand instructions, discriminate what is shown to it, select objects, and answer through the left hand; its organized knowledge and purposive responses support attributing phenomenality to the hemisphere-level center despite the loss of verbal report. The Wada procedure provides a reversible comparison by anesthetizing one hemisphere while the other is assessed (Loring et al., 1992), and hemispherectomy evidence confirms that hemisphere-level organization can sustain a conscious life (Blackmon, 2016). A disconnected right hemisphere is a biological instance of *reporter subtraction*: removing its route to speech eliminates testimony while preserving the knowledge state from which testimony could have been produced. Blackmon presses the further possibility that, even in an intact brain, each hemisphere sustains its own perspective while their interaction realizes a more

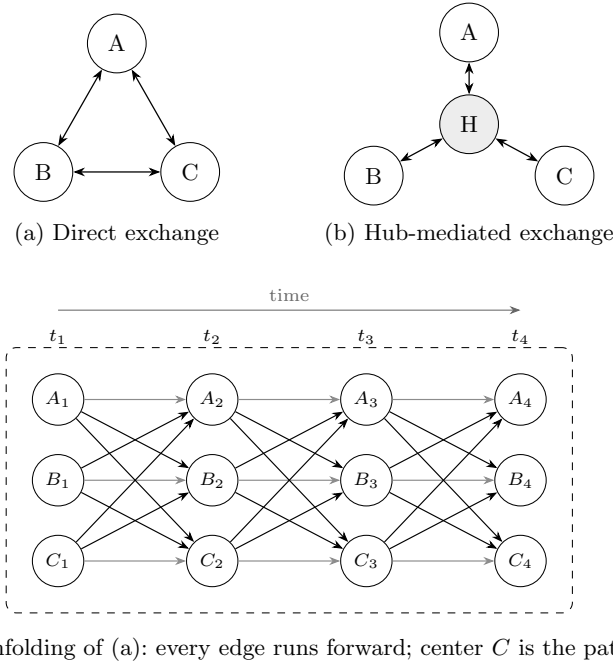


Fig. 2 Routes to broad access, and the center they realize. (a) Modules exchange selected outputs directly. (b) Modules exchange through an integrating hub. Both are component-level projections and both contain cycles. (c) The event-level unfolding of (a): each participation of a module at a stage is a distinct event, every edge runs forward in time, and the graph is acyclic. Grey edges are a module's own continuation; black edges carry distinctions between modules. The center is the dashed pattern of events, not any component within it. The hub architecture in (b) unfolds to a graph equivalent to (c) at the cognitive substitution level: its hub events are relay vertices that add path length without adding distinctions. Neither architecture contains a privileged locus at which experience must occur.

integrated whole-brain center. The first question to ask of any report from such a case is therefore: *which process is reporting?*

A state is commonly called “unconscious” when it is unavailable to the dominant center governing integrated report, episodic memory, and deliberate control. Two architectures produce that result: information may pass through activity that forms no center, or it may belong to another center whose knowledge is inaccessible to the dominant one. The autonomous vehicle contains both. A camera stream that no module reads passes through the vehicle without forming a center. The anti-lock braking controller is different: it possesses wheel-slip distinctions and uses them to modulate brake pressure many times a second, while the planner that commanded the braking never receives them and learns only that the vehicle stopped. It is a center of its own, nested inside the larger one, and what the planner's record would call unconscious is, for it, the whole of what it knows. Both architectures yield the report “I was unaware of it,” which establishes absence from the reporting center but cannot decide between the architectures. Chalmers raises the same possibility: processing unconscious to the person may support experiences belonging to subsystem-level subjects (Chalmers, 1996a, pp. 299–300). Learned skills illustrate the distinction: detailed motor control

begins with deliberation and becomes unavailable to report as practice consolidates it into specialized circuitry (Zuboff, 1994). Dream construction and unattended problem solving pose the same question: a finished image, sentence, or solution shows that information was processed outside the reporting center, and the intermediate work belonged to another center only if its discriminative states governed its own transitions. Whether such a center is phenomenal depends on the boundary question of Section 7. Block’s distinction between access and phenomenal consciousness remains useful when access means broad availability for reasoning and report (Block, 1995). Here access is indexed to a process, so locally centered knowledge determines a local center while broader access determines which process can report it. Information merely encoded or transmitted without entering a usable state remains outside centered knowledge altogether.

5.3 Nested, overlapping, and distributed centers

The access graph need not stop at the skull. Two uploaded minds could inhabit one computer and remain as separate as minds in different skulls, while a high-bandwidth bridge could join components on different continents. This is the **co-location principle**: physical inclusion and proximity do not create access, and spatial separation does not preclude it. Dneprov’s 1961 story *The Game* has fourteen hundred mathematicians in a stadium pass binary signals according to fixed rules, together implementing a Portuguese-to-Russian translation that none of them understands (Dneprov, 1961; Kirtchik, 2023). Dneprov treated their ignorance as evidence against machine thought, but the access graph supports the opposite reading, and as the Churchlands observe, “no neuron in my brain understands English, although my whole brain does” (Churchland & Churchland, 1990, p. 37). Block’s Chinese Nation applies the structure to consciousness (Block, 1978). Under organizational invariance, participants who reproduce the causal organization of a conscious brain collectively implement a conscious process. Under the **non-exclusion principle**, they retain their individual perspectives even as their interactions constitute a further center: a participant’s private thoughts lie below the larger process’s substitution level unless they alter the assigned signals, and no participant need have access to the integrated states of the population. Ant colonies and corporations extend the point beyond a constructed example, and Schwitzgebel has argued that a consistent materialist should take seriously that a sufficiently integrated nation is conscious (Hölldobler & Wilson, 2009; List & Pettit, 2011; Schwitzgebel, 2015). If such a center is phenomenal, its observer-moments may unfold over seconds, days, or years. This is the **boundary-independence principle**: the boundary of a center follows the process that constitutes it, not the boundary of any one object.

Non-exclusion has a cost that should be stated plainly. Because qualifying subgraphs need not partition G , a brain contains many overlapping centers, and under the minimum-boundary hypothesis of Section 7 many overlapping subjects. The account accepts this. Parsimony is a constraint on postulates, not on instances: a theory with one criterion and many centers is simpler than a theory with the same criterion and an added rule for suppressing all but one of them, and nature is not sparing with instances of anything else it produces. Integrated information theory’s exclusion postulate buys uniqueness at the price examined in Section 10.2.

6 A cumulative case for phenomenal identity

Why accept the identity rather than the weaker claim that centered knowledge is the functional correlate of experience? No definition of computation or knowledge forces a theory of phenomenality. The case is abductive. It is best stated against the strongest alternative: the *correlate view*, on which everything said so far about K_C is true, but K_C is the physical basis with which experience is lawfully aligned rather than experience itself. Each of the following considerations is compatible with the correlate view. What the correlate view must add, at each point, is a bridge that keeps the two structures aligned, and it is the bridge, not the correlation, that the causal dilemma of Section 4.3 then attacks.

The meta-problem and immediate knowledge.

Phenomenal judgments are causal outputs: a center can judge red rather than blue only if the corresponding distinction reaches the organization producing the judgment. And the center may know which state it occupies without knowing how it is implemented. Together these create an explanatory demand. A merely correlated state explains why a report occurs but not, by correlation alone, why the report is knowledge of experience. A property causally isolated from the judgment-producing organization supplies a purported object of knowledge but cannot explain how the judgment acquires contact with it. The identity proposes one relation that plays both roles: the distinctions in K_C supply the content of phenomenal judgment because they are the causally operative states from which judgment proceeds, and they are available to that judgment because the center occupies them, so that whether a given judgment is correct is a question about the competence of the organization that forms it and never about how the judgment reaches its object. The correlate view must instead say that the judgment is about the experience because the experience is lawfully aligned with the state that produces the judgment, which is to explain epistemic contact by a correlation that the subject cannot check.

Epistemic contact and replacement.

Shoemaker's point converts into a counterfactual test. Begin with a system whose consciousness is not in doubt and vary its implementation while preserving everything constitutive of K_C . If its experience changes in a way it can notice, compare, remember, or use, the organization was not preserved. If the alleged change affects none of these, it is detached from every route through which the center knows its experience. Kirk's dilemma generalizes: any phenomenal ingredient added to the organization either changes nothing the process can notice, and then cannot explain epistemic contact, or changes something, and is then a usable difference already within the complete organization (Kirk, 2008). A newly discovered variable revises the model of K_C , not the identity (Appendix A.3). The correlate view accommodates replacement by positing that the bridge law is organization-sensitive at exactly the substitution level; the identity predicts it.

Evolutionary alignment.

Natural selection tracks only heritable differences that affect outcomes. If pleasure and pain could be reversed while leaving the organism's causal organization unchanged, selection could not favor the actual pairing over its inversion. Cutter and Crummett use such inversions to expose an unexplained *psychophysical harmony* between experience and behavior (Cutter & Crummett, 2025), and Mørch's phenomenal-powers view answers that phenomenal properties necessitate their effects in virtue of how they feel: pain motivates avoidance because of what it is like (Mørch, 2017). That necessitation runs from phenomenal character to causal power and does not by itself establish the converse. The identity secures harmony in both directions. Pain is not an intrinsic quality that selection fastened to an independently evolving control state. Its character consists in its place within the center's organized distinctions—location, intensity, salience, aversive significance, memory, learning, available responses—and selection shapes phenomenal valence by shaping that organization, because the knowledge-bearing control state and the experienced state are one. In Resch's *Bots*, an artificial-life simulation, evolving neural-network controllers encountered green and red objects whose only assigned significance was their opposite effects on reproductive fitness, and selection produced pursuit of the beneficial objects and recoil and flight from the harmful ones, although neither policy was programmed (Resch, 2026). The simulation does not show that the controllers feel anything. It shows how selection turns a difference in consequences into an internal organization through which conditions matter to a process, which is what, on the identity, feeling would consist in. A phenomenal-powers theorist may hold that valence carries the motivational powers without accepting the converse, but then must say whether the fully organized powers can arise without phenomenality. If they can, they cannot explain why phenomenality accompanies them; if they cannot, the position approaches the identity.

Structural coherence.

Phenomenal colors form spaces of similarity, hue, saturation, and intensity; the visual field has spatial and object-relative structure; sounds vary through organized relations of pitch, timbre, and location. These relations closely match the distinctions available in perception, comparison, memory, and action. Chalmers calls this the principle of structural coherence, an isomorphism between the structures of consciousness and awareness, and because he holds experience and awareness to be distinct, the coherence remains for him a psychophysical principle to be taken for granted (Chalmers, 1995b, sec. 7). His proposed deeper explanation is a double-aspect theory of information, on which experience is information from the inside and physics information from the outside (Chalmers, 1996a, p. 305). The proposal comes close to the identity, but it takes information as Shannon does, as a pattern instantiated somewhere in a physical system, and a pattern so understood has no inside. The identity supplies the missing relation by locating information at a process. A usable distinction is information *for* a process: the same external pattern may reach two processes and place them in different states, because their transition organizations enact different interpretations, and what is information for one may be differently informative, or not informative at all, for the other. This is why one objective description of the world contains many distinct

and mutually inaccessible phenomenal aspects: each is the organization of distinctions some process possesses, interpreted through that process's own transitions, and a signal passed between centers gives the receiver a new input without transferring the sender's occupation of its state. On the correlate view structural coherence is a further principle to be taken for granted. On the identity, what appeared to be two correlated structures is one organization described from within and from without.

Unification and parsimony.

The same relation explains the results listed in Section 1.2: character, ownership, binding, causal role, invariance, privacy, and horizon. The identity would be merely verbal if centered knowledge were defined by reference to phenomenality, but it is not. K_C is specified independently through implemented distinctions, access relations, and counterfactual transitions, so the identity makes substantive commitments, one of which is forced rather than tested. Call it the **closure constraint**: given causal closure and the causal provenance of phenomenal judgments, every phenomenal difference a subject can notice, compare, or remember is a difference in the organization that produces the noticing. Any theory that respects the constraint predicts what the identity predicts wherever a subject could tell, and the identity is what remains when the constraint is respected and nothing beyond it is posited. What is tested is the other direction, that every usable distinction within a phenomenal center contributes a phenomenal contrast; a reproducible counterexample would refute it.

The rival accounts accept the processes that produce a center's discriminations, memories, and reports. Naturalistic dualism adds psychophysical principles; Russellian monism and phenomenal-powers views add non-differential intrinsic natures; substrate-sensitive views add particular physical powers; integrated-information theory adds a selection rule over physical complexes; strong illusionism subtracts the realist explanandum, and must then add an error theory for what it subtracts, whereas the identity's only error claim is the inference from occupation to a non-structural residue (Section 9.4). The comparative question is what work these additions do once the center's distinctions and the provenance of its judgments are fixed. To reject the identity, one must explain how centers with the same complete knowledge organization could differ in phenomenal status or character while the difference remains absent from every comparison, judgment, memory, and action through which either center could know its experience. Among the accounts considered, the centered account assumes the least while explaining the most: experience is what structured knowledge is like for the process that possesses it.

7 The phenomenal minimum and its boundary

The centered-knowledge thesis fixes phenomenal status and character across equivalent organizations but says nothing about which non-equivalent organizations are phenomenal. This section adds two further claims and says what each is worth: a claim that there is a boundary, and the identity's own answer to where it lies.

7.1 The phenomenal-minimum thesis

Boring wrote that “consciousness is, broadly speaking, discrimination” (Boring, 1933, p. 187), and James described it as a mode in which “data are brought together for comparison with a view to choice” (James, 1879, p. 8). These anticipate a lower-bound claim without locating it.

The phenomenal-minimum thesis. There is a minimally sufficient condition on complete centered-knowledge organization, $K_{\min}$, such that a center is phenomenal exactly when its K_C realizes it. The thesis leaves open what $K_{\min}$ requires.

K_C denotes the complete organization of a particular center; $K_{\min}$ denotes an organizational type, not a scalar quantity, a total ordering, or an amount of Shannon information. Global-workspace, higher-order, and recurrent-processing theories can be read as rival accounts of $K_{\min}$ insofar as their requirements are relations within centered organization: respectively, qualifying global availability, an appropriate higher-order representation, and recurrent access. A theorist may reject the phenomenal-minimum thesis while retaining the core thesis, by holding that the boundary is vague, that there are several incomparable routes to phenomenality, or that no least organization exists in a descending sequence. Such a theorist must still hold that exact organizational equivalents agree in phenomenal status and character. Appealing to substrate or biology to distinguish exact equivalents rejects the core thesis itself.

7.2 The subtraction argument

If locally usable possession were not yet experience, what further organization would make the possessed state experienced? Begin with one of the strongest cases: Yudkowsky describes a mind that catches its inner listener in the act of listening and says so aloud, and argues that whatever consciousness denotes participates causally in producing that report (Yudkowsky, 2008b). Now subtract.

First remove the capacity to speak. Disconnecting the output machinery eliminates testimony while leaving the reflexive awareness of listening intact; Wada and split-brain cases confirm that reportability and an organized point of view come apart (Section 5.2). Next remove the reflexive noticing while preserving the first-order listening state and its role in auditory processing. An actualist higher-order theory assigns phenomenality to the first-order state only when a suitable higher-order representation of it occurs (Rosenthal, 1986; Brown et al., 2019). The substantive question is why representation by a second state should create phenomenality in the first rather than add reflective knowledge about an already effective distinction, and Section 10.2 presses this. Then simplify the first-order content: move from a rich auditory scene to a barely discriminable tactile contrast, stimulation versus none at one location, preserving the state that registers the alternative and uses it while removing naming, recollection, evaluation, and self-description. Recurrence may sustain or elaborate the state, but reuse of a component is a fact about realization, not about the implemented process (Section 2.2, Figure 2c). Chalmers sketches a related subtraction: reportability is too demanding, introspection belongs to reflective rather than primary consciousness, sheer complexity supplies no threshold, and what remains plausibly essential is discrimination whose information is available for behavioral control (Chalmers, 1994).

Now subtract that remainder. If stimulation and nonstimulation leave the process in the same state, or every path from the selected state to its continuation is severed, the process possesses no knowledge of the contrast and, where this was its only usable distinction, no center remains. Earlier subtractions removed capacities while preserving a candidate centered organization; this one removes the relation required for any center to exist. It therefore supplies a natural lower limit for the inquiry. The intermediate steps show something weaker but useful: no proposed enrichment has been shown to belong to $K_{\min}$ unless removing it takes a process from some phenomenal perspective to none rather than from a richer organization to a sparser one. Integration explains binding, memory explains recollection, language explains report, and self-modeling explains reflection; none of these contributions shows that the capacity is required for phenomenality to begin.

A small mechanism makes the limiting candidate explicit. Let a sensor set a binary state q and a later stage use q to select a transition:

$$q_{t_1} = f(s_{t_0}), \quad z_{t_2} = g(q_{t_1}, z_{t_1}).$$

Two inputs select different q states, and g branches on q , as an **if** statement does, so that interventions on q that hold other conditions fixed alter z_{t_2} . Together these relations implement detection, propagation, and conditional use, and a trajectory lacks the third: a particle's later position depends on its momentum through one law applied alike to every value, so nothing in it does one thing for one class of momenta and another for another, and there is nothing on which its continuation branches. A barely detectable touch on the back of a hand is the human approximation: strip away tickle, temperature, pressure, and affect until only the faintest localized detection remains, and phenomenal character approaches the bare possession of a distinction, touched here rather than not.

7.3 The minimum-boundary hypothesis

The minimum-boundary hypothesis. A process realizes $K_{\min}$ as soon as its first implemented distinction becomes available to it and able to govern its continuation. No further multiplicity, integration, recurrence, global availability, or higher-order representation is required for phenomenality to begin.

Schematically, $K_{\emptyset} \xrightarrow{\text{first } \Delta^d} K_{\min}$, where $K_{\emptyset}$ is the absence of any centered organization at a locus rather than an existing center with empty knowledge. The first distinction has a capacity of one bit, but the boundary is not an amount: a bit no process branches on is a record, not a center. The hypothesis is the centered-knowledge thesis taken to its natural limit: it identifies the phenomenal minimum with the minimal center and adds no condition the center criterion does not already contain. Every alternative account of $K_{\min}$ adds a postulate, that some further organization is required before possessed and used knowledge is experienced, and the subtraction argument found no candidate whose removal takes a process from some perspective to none rather than from a richer perspective to a sparser one. The hypothesis is therefore the account's null hypothesis about the boundary, on the parsimony argument of

Section 5.3. A theorist who requires more owes an account of what the addition supplies, why phenomenal presence tracks the distinctions a center possesses rather than the addition, and how the center acquires epistemic contact with it, and must still explain why a center lacking it can possess causally operative knowledge with nothing it is like to occupy that locus. The core thesis does not entail the hypothesis, and the subtraction argument does not prove it. What would count against it is a genuine center with a nonempty, causally usable K_C and independent evidence that it lacks experience (Section 10.5).

What the hypothesis attributes experience to should be stated exactly, because the attribution is narrower than it sounds. A center requires a distinction that is implemented, that propagates far enough to enter the process's organization, and that is used to govern a continuation. A thermometer with no controller, a particle, a rock, a passive record, and most of the matter in the universe form no center at all, and the hypothesis says nothing about them. It says that every center is a phenomenal center, including the two-state mechanism of Section 7.2, a thermostat whose controller possesses and uses its distinctions, and the autonomous vehicle's anti-lock controller (Section 5.2). Chalmers asks whether a thermostat's three information states correspond to three phenomenal states (Chalmers, 1996a, pp. 293–297); the hypothesis answers that they do, and the core thesis then identifies the character of each with its sparse centered organization. Dennett sees no “magic moment” separating a thermostat's minimal representation from progressively richer ones (Dennett, 1987, pp. 31–33). The hypothesis agrees and adds that the continuum has a principled endpoint: the first center.

That a mind can be this simple is hard for intuition to accept, and an analogy helps to locate the difficulty. A gulf of complexity separates a bacterium from a whale, yet both are alive, no one supposes that there are two kinds of life or that life really begins somewhere between them, and nothing sets an upper limit on how complex a living thing can be. Conscious minds, on the present account, have the same shape: one relation throughout, gulfs of complexity between its instances, no upper bound, and a simplest case that consists in a single possessed distinction. The analogy has a second lesson. Life was once thought to require exactly what the Russellian thinks experience requires, an intrinsic extra over and above metabolism, reproduction, and heredity. When those were explained structurally no residue remained, and the question whether a bacterium is really alive or only functionally alive stopped being asked. The analogy is imperfect in one respect that favors the account: biology's borderline cases arise from the imprecision of its definitions of life, whereas the center criterion is stated precisely enough that a candidate either satisfies it or does not.

Nor is the simplicity of the minimal center a cost the rivals avoid. Each has a minimal instance of its own, and the rows of Section 10.1 are those instances: the mechanism of Section 7.2 with one register added that represents the sensor's state is the minimal higher-order system; three such modules reading one hub (Figure 2b) are the minimal workspace; a sensor and a memory element are the minimal integrated system, and Tononi and Koch grant such a circuit “a modicum of experience” whose only distinction is “this rather than not this” (Tononi & Koch, 2015, p. 11). The added components change nothing the system can do, or, where they are given work, are

usable distinctions already within the organization (the dilemma of Section 6), and each rival then faces a dilemma: stated without a threshold of size or richness, its condition is met by a system as small as ours and marks no onset that capacities reveal; stated with one, a thought with the right content, an ignition across a brain, enough integrated information, the threshold is a number with no principled location. The hypothesis takes the first horn and says why: conditions that do no work the capacities show are conditions parsimony removes.

Two consequences follow for the rest of the paper. Claims about phenomenal character, epistemic contact, and the phenomenal equivalence of organizational duplicates depend only on the core thesis. Claims about thermostats, bare feed-forward controllers, the autonomous vehicle’s controllers, and local sub-centers being phenomenal depend on the hypothesis. A reader who rejects the hypothesis should read every such attribution below as conditional on whatever account of $K_{\min}$ that reader accepts.

8 How a world becomes a point of view

Many puzzles about consciousness arise from failing to distinguish four relations that can come apart (Table 2).

Table 2 Four relations a system can bear to a state or process

Relation	Meaning
Description	A system contains information <i>about</i> a state or process without occupying the state or implementing the process.
Implementation	The system actually realizes that process.
Access	A selected process can use the state in memory, inference, report, or action.
Occupation	The state is the selected center’s own current organized state and constitutes its present perspective.

8.1 Uncentered worlds and centered states

Let W be a complete world described without selecting a viewpoint and G its total event-level causal organization. The criteria of Section 2.3 individuate centers within G . W says what exists and how its events are causally related. When a center C within it is phenomenal, the pair (W, C) says how some of that world is represented and experienced from that center. Moving from W to (W, C) adds nothing to the world’s causal organization but identifies one epistemic locus already realized within it. A complete world-description can specify every center and every state while selecting none of them as *this* center for a process. Possession supplies that selection, and it is not information from nowhere: the selecting relation is implemented by the state at C .

Nagel formulates the apparent remainder sharply. Even after the world has been completely described in objective terms, there seems to remain the fact that he is Thomas Nagel: that these experiences are his and this body is his body (Nagel, 1965).

On the present account the objective description need not omit Nagel or the organization that constitutes his experiences, and it may truly state that he occupies a particular centered state. What the description cannot do is make its reader the one who occupies that state. Being Thomas Nagel is occupying it, and that is a position rather than a further fact the description left out. Every phenomenal center is a first-person locus and none is globally privileged: from the standpoint of W all centers are represented, while from within each center some of the world's information is available there and able to guide that process. A center need not judge “I am C ”, which requires reflexive access and a self-model, whereas first-order occupation does not. Indexicals make the relation familiar: “I,” “here,” and “this” acquire referents from a speaker, place, or indicated object, and “this is red for me” relates a perceptual state to an access graph and a center (Lewis, 1979; Perry, 1979). Centering thus supplies a coordinate of possession rather than an additional ingredient or fact in the world.

8.2 The exclusivity of occupation

Possession selects, and selection excludes. The exclusivity of occupation stated in Section 3.5—to occupy one observer-moment is to occupy no other, so that no center can occupy them all—has three consequences for the cases that follow. An ideal reasoner who knows everything, if knowing includes occupying, is not a possible center: to occupy Mary's complete state is to exclude the omniscient state, and to occupy the omniscient state is to exclude Mary's. The complete description that fixes every truth about an experience cannot, for any reader, be that experience, not because a truth is missing but because occupation is not a truth to be learned. It is a position to be in. Nothing is absent from the total description of the world; there are only distinct and mutually exclusive occupations of the observer-moments it describes. A center that possessed every third-person fact would therefore still occupy one observer-moment of its own, an organization that could take many forms, each unlike any other, and its knowledge would include every truth about Mary's red state without including the occupation of it. What such a center lacks is not hidden from it; it is elsewhere. And what a description can transfer is therefore bounded in advance by the position of its receiver.

Nesting does not evade the result. A large center may contain a red-seeing sub-center and have access to its distinctions, and it then occupies a red distinction as part of its own complete organization. It does not thereby occupy the sub-center's state, because phenomenal character is complete organization, and red embedded in a small state is not red embedded in a vast one, though the relations that make it red rather than orange can be the same in both, and it is those relations that a comparison of shades across states compares. What the containing center knows by occupation is what red is like within a state like its own; what it cannot know by occupation is what it is like to be the smaller center. This is Nagel's limit (Section 8.4) stated as a consequence of the framework, and it is one of the two reasons the gap of Section 1.2 is explained rather than assumed.

8.3 Mary: description without implementation

A printout of a program specifies a computation without executing it. Mary's case turns on the same distinction. She knows every scientific description of color while living in a black-and-white environment (Jackson, 1982; 1986). Reading those descriptions exercises linguistic and theoretical machinery, while seeing a red tomato exercises visual machinery. The two forms of information concern the same perceptual process, but they enter through different routes and create different states. Dennett contrasts learning about wine from a label with tasting it (Dennett, 1991, p. 384): label and taste concern the same wine but place the learner in different knowledge states. Mary likewise comes to know from within what her descriptions could only specify from outside, because she now occupies the red-seeing state. On the centered-knowledge account, what she learns is what red-seeing is like for a center occupying that state.

This knowledge is incommunicable in exactly the sense implied by centered access. Mary can describe her new abilities in third-person terms, since she can now recognize, imagine, remember, compare, and name red, and she can describe the organization that realizes them. No description transfers her present occupation of that state to a listener who remains in another state. Communication produces knowledge *about* her experience, whereas knowing it from within requires the receiving process to occupy a corresponding state. The case therefore requires no additional nonphysical fact. It requires only the distinction between describing a process and implementing it, and between representing a state and occupying it. Conee's acquaintance reply captures part of this (Conee, 1994), and the present account naturalizes it: Mary's new relation is not a primitive acquaintance with a further fact but the implementation and occupation of an organization her descriptions represented without instantiating. A sufficiently reconfigurable Mary could use a description to implement the red-seeing process within herself. Until she does, she has a description of a process she does not implement.

8.4 Nagel: description without occupation

Nagel connects consciousness to what it is like *for an organism* and observes that imagination remains bounded by one's viewpoint: imagining flight, echolocation, eating insects, and hanging upside down by one's feet tells him only what it would be like for a human to behave as a bat (Nagel, 1974). Imagination rearranges the resources of his present mind and does not supply the bat's centered state. To know the bat's state from within is to occupy a state with the bat's organization, and reconfiguring Nagel's process into that organization would produce an observer-moment with bat-like knowledge rather than a human one that comprehends what it is like to occupy the bat's state. That is the gap in its barest form: a process occupying K_N cannot at the same time occupy the incompatible K_B (the exclusivity of occupation, Sections 3.5 and 8.2). Mary and Nagel expose the same distinction: describing an experience does not place the describer in the state that constitutes it.

8.5 Searle: access without occupation

In the Chinese Room, Searle imagines himself following rules that transform Chinese symbols fluently enough to pass for a speaker, while understanding none of them, and

concludes that executing a program does not produce understanding (Searle, 1980). His reply to the systems objection sharpens the case: let him memorize every rule and run the whole procedure in his head, and he still understands no Chinese. That internalized version is the one the framework addresses, and in it Searle resembles a host computer running a guest virtual machine. The host executes its own instruction set and need not interpret the guest’s instructions as the guest does, even though it can inspect every symbol and intermediate state. Access to an implementation and occupation of the state it realizes are different relations. Cole makes the difference concrete by asking the Chinese-speaking process about its height, breakfast, or attitude toward Mao: its answers need not match Searle’s, and replacing Searle with another rule follower leaves the Chinese process with the same memories and opinions (Cole, 1991). Searle is a replaceable component of the implementation, whereas the Chinese speaker is the process whose organization Searle sustains. His sincere report that he understands nothing establishes that Chinese meanings are unavailable to him through ordinary introspection. It does not inventory every process implemented by his activity. The split brain is the biological parallel: the speaking hemisphere’s sincere denial does not settle what another center knows. The anti-lock controller of Section 5.2 is the mechanical one: even if they shared computer hardware, it and the planner are different centers.

8.6 Privacy

Privacy follows from access topology. Bob’s pain can cause speech or neural activity that Alice measures, but these create states in Alice *about* Bob and do not place her in his pain process. The autonomous vehicle is in the same position with respect to the vehicles it watches brake at a red light: their braking tells it that they registered the light, and even the full schematics of their controllers would give it a description of their red-selective states rather than the occupation of one. Proximity changes nothing by itself. The Hogan twins, whose separate brains are joined by a thalamic bridge, are reported to share some visual information that would ordinarily remain private (Cochrane, 2021). The evidence does not decide whether this is one shared experience or two closely coupled ones, and what would decide it is the contrast privacy turns on: receiving evidence about a state, which transmission supplies, or integrating that state’s constitutive organization into one’s own processing, which alone makes an experience shared. Locke’s worry that a violet might produce in one person the idea a marigold produces in another while both use “blue” and “yellow” normally (Locke, 1689/1975) is real to the extent that two centers’ color organizations differ in ways coarse tests miss, and empty to the extent that a proposed inversion preserves every centered relation, since covariance then entails that character is preserved too.

8.7 Blockhead and causal provenance

Despite privacy, we gain strong evidence of other minds. A sensible answer occupies a tiny part of the space of possible replies, and apt replies to novel follow-up questions test an expanding tree of counterfactuals, which is why confidence rises quickly after relatively few exchanged bits (Aaronson, 2007). Block’s Blockhead is the standing

objection: a lookup table containing an appropriate response for every possible conversational history would reproduce a mind's outward behavior without its internal organization (Block, 1981). Yudkowsky asks where such a table came from (Yudkowsky, 2008a). If its entries were produced by executing a conscious process, the transcript is recorded evidence of that process, and interacting with the table resembles hearing a voice on a recording: the transmitting mechanism need not possess the mind its output evidences, and that mind may be remote in time as a telephone voice is remote in space. Three things must therefore be distinguished: the stored table, the executing lookup process, and the earlier process that generated the entries. The lookup process discriminates addresses and uses the selected entry, so it forms a center of its own, but the beliefs, memories, and understanding its answers express belong to it only if its present organization realizes them, and otherwise belong to the process that produced the table. A table stipulated to arise by chance remains logically possible, and imagining one implies no earlier speaker, but it also gives no reason to think the behavior was ever produced by anything that knew what it was saying. This is the **causal-provenance principle**: an output can carry evidence of a remote mind even when the process presenting it does not itself instantiate that mind.

9 The observer's epistemic horizon

The distinction between occupation and description marks not only the boundary between minds but the boundary between an observer's present knowledge and the world, history, program, and hardware it infers from that knowledge. The organization that gives a center a world also fixes what it can know about that world.

9.1 The present-state boundary

Classical skeptical scenarios allow different causes to converge on the same observer-state: Descartes's deceiver, Putnam's brain in a vat, Russell's universe created five minutes ago with its records and memories in place, Boltzmann brains, and simulation hypotheses (Descartes, 1641/1984; Putnam, 1981; Russell, 1921; Bostrom, 2003). An embodied brain, a stimulated brain, and a simulated mind can occupy the same accessible state. Let $\mathcal{W}(K_C)$ be the set of worlds, in the sense of Section 8.1, together with their histories and hardware, that are compatible with the complete knowledge state presently available to C . That state underdetermines the world whenever $\mathcal{W}(K_C)$ contains more than one member, and no inference whose only input is that state can then uniquely determine which is true. A larger metatheory, prior, or measure may assign the worlds different probabilities, and a later test may separate them by adding evidence. An emulation that supplies the same enlarged history, however, preserves the underdetermination.

Present-state boundary. A process directly possesses only the distinctions in its current accessible state. Its beliefs about the world, the past, other minds, and its own implementation are inferences from that state.

This is not a license for global skepticism. Stable regularities across many transitions give powerful evidence for objects, histories, and agents, and simplicity favors some

hypotheses over others. The boundary says only that the inference must proceed from what is accessible now, and that what is accessible now is fixed by the center’s own organization.

9.2 Program underdetermination

Even if the external world and the hardware substrate were known, K_C alone could not identify a unique program. Three forms can be distinguished. First, K_C need not exhaust the complete instantaneous configuration of a running program: other event states in G , and lower-level variables realizing them, may differ without contributing to the current K_C . Second, the history available to C does not fix the transition rules: two programs may have generated every observer-state so far yet respond differently to an input not yet received, and further observation can exclude some rules while leaving others that diverge later. Third, fixing center-level transitions does not fix the lower-level routines that implement them: two libraries may compute the same function by different algorithms, and if both preserve every distinction and transition available at the substitution level their differences remain outside K_C . Marchal derives a related first-person indeterminacy from the observation that every program compatible with an observer’s history so far, including those that later diverge, is realized somewhere in the space of computations (Marchal, 2013). Introspection and experiment can narrow the compatible programs but cannot reduce them to one.

The horizon also bounds what one center can know of another. When a center’s next transition is computationally irreducible, no shortcut yields it: a predictor must carry out the transition itself (Wolfram, 2002; Israeli & Goldenfeld, 2004), and a model that does so at the center’s substitution level reimplements the observer-state it was meant to foresee. By organizational invariance that reimplementations is a second instance of the center, phenomenal with the same character. Prediction of a mind at its substitution level is therefore not a view of it from outside but the instantiation of another one, and what was to be prediction has become observation of the transition occurring there.

9.3 Universal emulation and ontological underdetermination

A universal Turing machine emulates any other machine given its description, $U(\langle A, p \rangle) = A(p)$, and Turing described universality as a digital computer’s ability to mimic any discrete-state machine (Turing, 1937; 1950). Combined with computational functionalism, universality implies that direct execution and faithful emulation preserve every transition available to C , including every internal test of its implementation. The process may run on native hardware, nested virtual machines, or a computer in another universe; the center can infer properties of the interface it receives while the ultimate realization of that interface remains underdetermined. Every finite segment of a computable observer-history can therefore be implemented in any world that permits universal computation, given suitable resources and initial conditions. This is **ontological underdetermination**: the accessible observer-state is compatible with many programs, substrates, histories, and underlying laws, and dreams, deceivers, vats, and simulations are indistinguishable to the center throughout any observer-history they preserve. Skepticism about substrate is thus not an exotic possibility grafted

onto computationalism but follows from universality together with the present-state boundary.

Nor does any of this depend on the physics being like that of our universe. Conway's Game of Life, with two rules on a grid, supports a universal Turing machine (Rendell, 2016). Under computational functionalism this settles a question about where consciousness could reside in physics: phenomenal character cannot depend on any special feature of our physical laws, discovered or undiscovered, because a world whose entire physics is two fully specified local rules can implement every organization that the identity says constitutes experience, and can therefore host minds as rich as any here. Only noncomputable physics escapes, at the cost noted in Section 10.3.

9.4 The horizon of the intrinsic

The underdetermination results have a positive form, and it concerns not what a center cannot decide but what it cannot in principle be given. A center possesses distinctions at its substitution level and nothing beneath. Whatever realizes those distinctions can be replaced by anything that preserves the same transitions, and the replacement is invisible from inside. It follows that, for any center in any world, the physics available from within will describe only structure and dynamics: which alternatives there are, how they are related, and how they change. This is not the claim that implementation is hidden. A center can extend its interface, add instruments, and learn how its distinctions are realized, and each extension delivers more structure and dynamics. What the relata are in themselves, beneath every structure any interface could deliver, is not a further fact the center has so far failed to discover. It is a fact no center could discover, because only distinctions and transitions can cross a point of view's boundary.

This explains a feature of physical theory that has been read as a discovery about matter. Galileo and Locke separated the geometrical and dynamical properties of bodies from the qualities that arise in a sensing body (Galilei, 1623/1957; Locke, 1689/1975); Russell and Eddington observed that even the primary qualities are known only through their relations, so that physics tells us the structure of the external world and nothing of its intrinsic character (Russell, 1927; Eddington, 1928); and Chalmers's structure-and-dynamics argument is the contemporary statement of the same observation (Chalmers, 2003a). Eddington added that there is one case in which we know the intrinsic character directly, because we are it: our own experience. Together these observations have seemed to leave a vacancy in the physical description exactly the size of consciousness, and Russellian monism fills it by identifying the unknown intrinsic natures with the one intrinsic nature we know.

The centered account grants the first observation, which is also the first premise of Chalmers's argument, and derives it: the structural character of physics is a consequence of the account of points of view. It disputes that argument's second premise, that phenomenal truths are not structural, as Section 4.3 argued. And it reinterprets Eddington's second observation. What Eddington had from the inside was not a glimpse of a categorical base but the one case in which a subject occupies a structure rather than describes it, and the felt intrinsicness of experience is what occupation is like from the position of the center (Section 3.5): possessed without inference, unavailable through description, specified by more relations than can be verbalized. Both of Eddington's

data are structural. Two things remain unknown from any point of view, what realizes the relata and what it is like to be a center one does not occupy (Section 8.2), but they are unknown for different reasons, and there is no ground for placing one mystery inside the other merely because both are mysteries. Section 10.3 draws the consequence for Russellian monism. The attraction of Russellian monism is therefore explained rather than shared. Physics looks like a structure with something missing because every observer sees only structure, and experience feels like the missing intrinsic ingredient because occupation is what structure is like from inside. Once both appearances are accounted for by the horizon, the vacancy the view was built to fill dissolves. This is the sense in which the boundary that individuates a center is also its epistemic horizon: what reaches a center through its access paths is all it can know, and what its relata are beneath all structure is what it can never know.

10 Rivals, predictions, and tests

This section compares the account’s attributions with those of its main rivals (Section 10.1), takes up the two sharpest disagreements, with higher-order theories and integrated information theory (Section 10.2), states the challenge the dilemma of Section 4.3 poses for nonfunctionalist theories (Section 10.3), and proposes tests of the core thesis (Section 10.4) and of the boundary (Section 10.5).

10.1 Contrasting attributions

Table 3 makes the disagreements with three rivals explicit. The global neuronal workspace (GNW) column treats recurrent amplification and broad broadcast as the criterion for conscious access (Mashour et al., 2020); the higher-order thought (HOT) column represents the actualist view requiring an appropriate representation of the current first-order state (Rosenthal, 1986; Brown et al., 2019), and attention schema theory, on which awareness is the content of a simplified model of the system’s own attention (Graziano, 2013; Webb & Graziano, 2015), shares that column’s verdicts except in the third row, where a representation of the first-order state is not yet a model of attention; the integrated information theory (IIT) column attributes experience according to intrinsic physical cause–effect structure and selects complexes by maximal irreducibility (Albantakis et al., 2023). In the centered column, *core* marks verdicts that follow from the identity because the target preserves the complete organization of an acknowledged phenomenal center; *boundary* marks verdicts that depend on the minimum-boundary hypothesis or another accepted account of $K_{\min}$. Where the stipulated equivalence fixes every report, no behavioral experiment can separate the theories, and the contrast is a test of which attribution is defensible.

The process in the first row is a center in the demanding sense of Section 2.3, with a distinction implemented, propagated, and used; a thermometer with no controller or a particle would not qualify and receives no attribution from any column. Every disagreement between the centered column and GNW occurs in a boundary row; in the core rows the two agree. The identity as such therefore does not compete with GNW; the boundary hypothesis does, by treating broadcast as the organization of a large, richly accessible center rather than the onset of phenomenality. Treating absence from

Table 3 Consciousness attributions in specified architectures. “Conscious” means that the theory attributes at least one phenomenal distinction to the target; “underspecified” means the stated architecture does not determine a verdict without further facts.

Architecture and target	Centered	GNW	HOT	IIT
Discrimination and use. A feed-forward component architecture possesses and conditionally uses a discriminative state; no broadcast or higher-order representation. <i>Target: the process.</i>	Conscious (<i>boundary</i>)	Not conscious	Not conscious	Not conscious
Global broadcast. First-order discrimination and workspace broadcast; no qualifying higher-order state. <i>Target: the first-order content.</i>	Conscious (<i>boundary</i>)	Conscious	Not conscious	Underspecified
Higher-order representation. First-order state and higher-order state integrated in one center; no broadcast. <i>Target: the first-order content.</i>	Conscious (<i>boundary</i>)	Not conscious	Conscious	Underspecified
Detached monitor. A first-order state is broadcast within one center; a higher-order representation of it exists only in a separate center with no return path. <i>Target: the first center’s state.</i>	Conscious (<i>boundary</i>)	Conscious	Underspecified	Underspecified
Faithful emulation. An acknowledged phenomenal center’s complete computational organization is preserved, including broadcast and higher-order state; physical cause–effect structure changes. <i>Target: the emulated center.</i>	Conscious (<i>core</i>)	Conscious	Conscious	Next to nothing
Overlapping complex. An acknowledged phenomenal component retains its complete organization while embedded in a larger overlapping complex with greater integrated information. <i>Target: the component.</i>	Conscious (<i>core</i>)	Conscious	Conscious	Not conscious

the dominant workspace as absence everywhere repeats the access error exposed by the Chinese Room and the divided brain. The higher-order column’s “underspecified” in the detached-monitor row records a choice the theory has not made: whether the higher-order representation must affect, be accessible to, or belong to the same center as the first-order state. The disagreements with HOT and IIT are sharper and are taken up next.

10.2 Higher-order theories and integrated information

The idle-monitor toggle.

Retain the first-order mechanism of Section 7.2 and add only a monitor variable m_{t_2} that equals 1 if q_{t_1} is in the monitored state and 0 otherwise. Stipulate that m has no outgoing causal role: no later transition equation contains it. Deleting or toggling m leaves every trajectory of q and z unchanged and alters no discrimination, action, memory, or report. If a lean actualist HOT account nevertheless makes q unconscious

without the suitable value of m and conscious with it, it assigns constitutive work to a downstream-idle variable. Now imagine a malfunction that alternately enables and suppresses m while leaving q and every downstream transition of a person-level process fixed. On the lean account consciousness would blink in and out, yet the person could show no confusion, attempt no compensation, report no interruption, and form no memory of one. This is a higher-order version of dancing qualia (Chalmers, 1995a). By the anti-zombie principle stated in Section 10.4, the acknowledged phenomenal person remains phenomenal if toggling the idle monitor preserves the complete relevant organization, and the identity predicts unchanged character. Avoiding the result requires giving the higher-order state a causal role or making it available to the same center—but either move turns the monitor into a further usable distinction already represented in K_C , which is exactly the centered account’s treatment of higher-order representation: it enriches and reorganizes a point of view rather than creating one.

The argument has a definite target and a definite limit. Rosenthal has held that consciousness makes little functional difference and that a higher-order thought need not affect the first-order state it is about (Rosenthal, 2008), and for that view the monitor can be idle, so the toggle bites. Contemporary higher-order theories that give the higher-order state metacognitive work—perceptual reality monitoring, on which the higher-order state gates whether a first-order signal is treated as reflecting the world, or views on which the higher-order representation is itself the phenomenal state (Lau, 2022; Brown, 2015)—are not touched by the toggle, because their monitor is not idle. Attention schema theory belongs with them, since its model of attention does causal work in control and report. Against them the disagreement is the boundary question, whether a first-order center lacking any such monitor is already phenomenal, which the toggle cannot decide and the tests of Section 10.5 address.

Integrated information.

Tononi and Koch argue that to exist from its own intrinsic perspective a system “must have cause–effect power upon itself” (Tononi & Koch, 2015, p. 7). IIT and the centered account agree that a distinction counts for a system only when it can make a difference to that system’s own history, which is Bateson’s difference that makes a difference indexed to the system itself. The two accounts diverge on three further IIT commitments. First, IIT identifies the conscious substrate as a set of physical units at a definite grain and selects a maximally irreducible complex, so that a faithful digital emulation receives, in Tononi and Koch’s words, “next to nothing,” because “the brain is real, but a simulation of a brain is virtual” (Tononi & Koch, 2015, pp. 1, 15). Second, feed-forward systems have no irreducible cause–effect power and are unconscious regardless of what they compute. Third, the exclusion postulate lets a maximal complex supersede overlapping component minds.

Against the first two, the event-level account gives a precise form to the unfolding argument (Doerig et al., 2019). At the level of physical events, causal influence always passes from an event to a later one; a feedback path reaches a later state of a component, never the event from which it began. Temporal reuse of one component and spatial replication of its successive uses realize equivalent labeled event graphs when

they preserve the same states, dependencies, and counterfactual transitions. Different phenomenal attributions would then track the assignment of events to physical components rather than anything available to the process. IIT must either identify a cause–effect power omitted from the time-indexed account or explain why grouping successive events under one persisting component determines experience when that grouping adds no causal relation to the center’s history. Chalmers raises a related meta-problem challenge: a high- Φ system and a computationally isomorphic low- Φ simulation generate the same reports and intuitions, so integrated information plays no role in explaining them (Chalmers, 2018, p. 37). The centered account strengthens the challenge under a condition more demanding than behavioral equivalence: if the emulation preserves the complete centered organization, it preserves the states that give the reports their content and provenance, and assigning different phenomenal status creates a difference the affected center could never notice.

The unfolding argument has drawn replies that should be met. Tsuchiya, Andrillon, and Haun object that it presupposes functionalism by treating input–output equivalence as the relevant standard (Tsuchiya et al., 2020). The event-level version does not: it requires equivalence of the complete internal event graph, including counterfactual transitions, and asks only that a theory say what causal fact the recurrent realization possesses that the unfolded one lacks. Kleiner and Hoel argue that the argument generalizes: any theory whose predictions can dissociate from a system’s reports is empirically unfalsifiable, and functionalist theories face a mirror-image substitution problem (Kleiner & Hoel, 2021). The centered account accepts the first half as a constraint it satisfies by construction—its predictions cannot dissociate from what the center can notice—and accepts the second as the price of that constraint (the closure constraint of Section 6): a substitution that preserves the complete K_C cannot change what the subject can report, so substitution tests locate the substitution level, and the account is exposed instead through the other direction of covariance, a usable distinction within an acknowledged phenomenal center that contributes no phenomenal contrast (Appendix A.3).

Against exclusion, consider a worker whose job is to serve as neuron 283,782 in a China Brain, receiving messages from assigned coworkers and sending the prescribed signals onward (Block, 1978). If the organized workforce becomes a more integrated complex than any constituent brain, IIT attributes consciousness to the China Brain while excluding the workers’ overlapping brains. The worker becomes a zombie on clocking in: for eight hours she converses, remembers her family, and sincerely insists that she is conscious, while no experience belongs to the person producing those words. A sufficiently disconnected lunch break restores her consciousness, and clocking out restores it for the evening. Nothing in her brain is destroyed or rebuilt, yet her phenomenal status turns solely on whether her intact brain is currently embedded within a more irreducible complex. Goff draws the same consequence for a sufficiently integrated society (Goff, 2019). The centered account permits the workforce to constitute an additional center without extinguishing its component minds. Recent adversarial testing of IIT against GNW yielded results compatible with parts of both while challenging key predictions of each (Cogitate Consortium et al., 2025). Broadcast and integration are empirically tractable ways of structuring which distinctions become jointly available

within a center, and evidence for either supports the centered description of rich organization without showing that either is necessary for phenomenality.

10.3 Challenges for nonfunctionalist theories

The dilemma of Section 4.3 applies to any proposed consciousness-making property X beyond the center's causal organization: if changing X alters anything available to the center, the functional description was incomplete and an adequate replacement must preserve its effects; if it alters nothing, it cannot influence what the center notices, remembers, compares, judges, or reports, and a stipulated phenomenal difference is disconnected from every route by which the process could know it.

Substrate intrinsicism and biological naturalism. Consider three computers implementing the same process in silicon, germanium, and carbon nanotubes, with every distinction, transition, access relation, and timing dependency preserved. Their materials and microscopic histories differ, but nothing in the implemented process unfolds differently. To assign them different experiences because their materials differ would make character vary without any difference the processes could encounter. Biological naturalism holds that brains possess causal powers absent from programs (Searle, 1980). If those powers are physically efficacious, their effects appear in the distinctions and transitions available to the center and belong to the organization an adequate implementation must preserve, if necessary by simulating the physics that realizes them; if they produce no such difference, they cannot explain why the biological process feels different. Only a causally effective, noncomputable influence absent from present physics escapes this dilemma, and that abandons computational functionalism.

Non-algorithmic physics. Penrose's case begins with a claimed behavioral difference: human mathematicians possess capacities no algorithm can reproduce (Penrose, 1989; 1994). Gödel's theorem does not establish this: it shows that a sufficiently expressive consistent formal system leaves some truths unproved within that system, not that fallible mathematicians transcend every algorithm (Feferman, 1996; LaForte et al., 1998; Aaronson, 2007). Even if such a capacity were established, Orch OR would not follow: it requires an unconfirmed gravity-induced reduction supplying noncomputable dynamics, since ordinary quantum computation changes resource requirements without extending the computable functions (Deutsch, 1985; Bernstein & Vazirani, 1997), microtubules that sustain and orchestrate the relevant states, and an account of why those reductions constitute a point of view (Hameroff & Penrose, 2014). Each step needs independent evidence, and the constitutive step faces the general test: if objective reduction changes judgment, its effects enter the center's organization; if it changes nothing there, it explains neither the alleged mathematical advantage nor the process's experience. Field theories of unity face the same test (McFadden, 2020).

Russellian monism. Section 4.3 conceded that a view on which phenomenal or protophenomenal properties are the categorical grounds of causal powers escapes the causal-exclusion horn of the dilemma, since on that view the grounds are not an addition to the structure but what has the powers. It does not escape the epistemic horn. A structural duplicate with different or absent grounds, and therefore with different or absent phenomenality, makes every judgment the original makes, so no judgment tracks the grounds, and acquaintance with one's own intrinsic nature is

the paradox of phenomenal judgment under another name (Russell, 1927; Chalmers, 2015; Alter & Nagasawa, 2015). The centered account and Russellian monism are nevertheless the same kind of theory. Both are dual-description monisms: one thing, described structurally by physics and as experience from within. They differ only in where they put experience. The centered account puts it in the relational organization of a center; the Russellian puts it beneath the relations, in what the relata are. That difference decides everything that follows. Relational organization is differential: it makes differences to judgment, memory, and action, so a center can be in contact with it, and it has parts and relations, so it can explain why an experience has the structure it has. Intrinsic nature is non-differential and non-relational by definition. Placing experience there is therefore an explanatory dead end in three respects. No experiment, report, or theory can distinguish one assignment of intrinsic natures from another, since every such act is structural; the view cannot be improved by any evidence. It cannot explain phenomenal structure, since a non-relational nature has no structure to explain with, so the qualitative geometry of experience must be borrowed back from the relations it was meant to ground. And it relocates the explanandum rather than discharging it: why this structure feels as it does returns as why this nature feels as it does, with fewer resources for an answer. Parallelism and idealism ended in the same place for the same reason, at a level said to underwrite everything and open to interrogation about nothing. The phenomenal-powers view of Section 6 is the one attempt to make the grounds differential; it was answered there. Section 9.4 explained why the Russellian identification is nonetheless tempting. Whether structure needs categorical grounds at all can be left to the metaphysics of physics (Stoljar, 2001). On either answer they are idle with respect to which centers exist, what they know, and what their experience is like.

Panpsychism. Constitutive panpsychism assigns experience to the fundamental relata directly (Goff, 2017). In its Russellian form it inherits the idleness just described and adds the combination problem: it owes an account of how micro-subjects compose a macro-subject with one perspective, while the centered account already has one, since composition is joint usability in a common continuation (Section 5.1). The panpsychist therefore needs the centered account for everything a subject could know about its experience, and the centered account needs nothing from panpsychism. In its non-Russellian form, on which micro-experience is an addition to physical structure rather than its ground, a twin-world argument applies directly. Suppose a world preserved every particle trajectory, interaction, and causal relation while lacking the microphenomenal properties posited in ours. The same history would unfold, including philosophers writing word-for-word defenses of panpsychism and crediting properties their world lacks. Those properties therefore explain neither the judgments nor the knowledge offered in their support, and nothing in the physics of everyday matter leaves room for them to do otherwise (Carroll, 2021). The Russellian panpsychist will deny that such a twin world is possible, and for her the argument of the preceding paragraph applies instead. The two forms fail for different reasons, but both leave the theory of consciousness where the centered account already stands.

10.4 Testing the core thesis

Two premises, each stated independently of the identity, fix what a positive result can establish.

- P1. Phenomenal-judgment premise.** An entity that originates qualifying judgments about its own experience is conscious. A judgment qualifies when it arises from the entity’s own relevant states, responds appropriately to interventions on them, and does not merely repeat a supplied answer.
- P2. Anti-zombie principle.** If one entity is phenomenal, an organizationally equivalent entity cannot lack phenomenality, where equivalence concerns the complete relevant causal organization, including counterfactual dispositions, rather than outward behavior.

P1 is motivated by the difficulty of treating independently generated, intervention-sensitive judgments about one’s own experiential states as systematic errors in an entity that has no experience (Yudkowsky, 2008b; Argonov, 2014). An illusionist will reject it, and results that depend on it are marked below. P2 is not argued for here. It states what a reader who denies the possibility of zombies is committed to, and results that depend on it are marked below, as those depending on P1 are. Writing $J(A)$ for “ A originates qualifying phenomenal judgments” and $A \simeq B$ for organizational equivalence: $J(A) \Rightarrow A$ is conscious, and $(A \text{ phenomenal} \wedge A \simeq B) \Rightarrow B$ is phenomenal.

The following tests begin with independently acknowledged phenomenal systems or, in the provenance case, with a judgment-producing entity whose phenomenality follows conditionally from P1. None infers phenomenality from discrimination and use alone. The tests bear on empirical models of K_C ; the identity itself is assessed abductively in Section 6, and a result counts against it only through the models it licenses. A reproducible difference in phenomenal status or character with complete centered organization preserved, or a usable distinction within an acknowledged phenomenal center that contributes no phenomenal contrast, would challenge the identity.

Recurrence-necessity comparison.

Replace a recurrent, experience-related circuit in an artificial mind, or eventually an uploaded one, with its feed-forward unfolding while preserving every causal relation at its interface. An autoregressive language model illustrates the equivalence: generating 1,000 tokens by reusing one machine 1,000 times or by passing the complete inference state through 1,000 machines with identical weights realizes equivalent event graphs, and reusing the hardware is economical rather than constitutive of the computation. The horizon clause of Section 2.3 matters here: at each stage the two realizations have equivalent K_C because the present state governs the same next transitions through the same interface, and each further stage costs the one temporal depth and the other circuit depth. If the substitution changed any experience, the subject could notice or report the change only if some later transition also changed, and the unfolding preserves those transitions. A theory that makes physical recurrence constitutive must posit a phenomenal change to which the subject has no access; repeated switching

turns the case into recurrence-specific dancing qualia. Transformers already weaken the motivation for treating recurrence as the threshold, since systems built from feed-forward passes perform translation, reasoning, and image interpretation without recurrent connections within a pass (Vaswani et al., 2017). This does not show that such systems are conscious (Chalmers, 2023; Butlin et al., 2023), only that recurrent circuitry is not required for rich discrimination or workspace-like availability, and an interpretability study of feed-forward language models has since identified a sparse set of verbalizable representations with the broadcast, routing, and selectivity properties associated with a global workspace (Gurnee et al., 2026). Because the stipulated equivalence fixes every report, this is a conditional argument about attribution rather than a behavioral experiment. It challenges recurrence requirements under P2 without assuming the identity, and the identity adds the prediction of unchanged character.

Neural-substitution test.

Anchor a person-level center C_R on the language-reporting event-subgraph. Structural connectivity supplies a component-level projection of possible routes into it; time-resolved effective connectivity and direct interventions determine which event-to-event dependencies actually carry distinctions into comparison, memory, report, and action. For a visual or auditory pathway, locate the coarsest interface whose exported distinctions, timing, and counterfactual effects suffice to preserve K_{C_R} : this is the substitution level the identity predicts for that process. Replace machinery beneath the interface with prostheses of progressively coarser grain. The identity predicts that every replacement preserving the interface leaves the participant’s perceptual discriminations, comparisons, and reports unchanged, and that once coarsening omits a constitutive relation some available distinction changes. Within-subject cases are strongest. A participant who developed with sight or hearing and later lost it can compare prosthetically restored states with remembered ones across color, pitch, location, similarity, and binding, though the comparison then runs through memory. A participant whose loss is confined to one side can do better: the restored half of the visual field can be compared with the intact half at the same moment, as a cochlear implant in one ear is already compared with normal hearing in the other, and the comparison no longer depends on memory. This is Zuboff’s hemispheric replacement (Section 4.1) as an experiment. A stable phenomenal difference produced by a replacement below the predicted boundary would falsify that boundary, a coarser replacement that preserved every discrimination would show that the predicted level included details that were not constitutive, and a reported difference under a replacement that preserved everything at the predicted interface would show that the interface omitted a relation, since the report is itself a transition the organization governs (the closure constraint of Section 6). Matching reports are evidence about fidelity, not proof of complete equivalence.

The horizon of Section 9.4 has a practical corollary here. A center cannot certify from within that a reconstruction preserved its phenomenal character, because the standard of comparison is rebuilt along with everything else: Sacks’s colorblind painter lost not only color vision but the ability to remember, imagine, and dream in color (Sacks, 1995). A reconstructed mind’s own report that nothing has changed is therefore not self-certifying, and absent a theory that fixes the level, the level will be set by

guesswork. Sandberg and Bostrom grade emulation success from a parts list up through behavioral fit to mind emulation and personal identity, and note that the higher criteria cannot be checked from outside (Sandberg & Bostrom, 2008). Reports are the wrong instrument for that check. The human optic nerve carries of order 10^7 bits per second (Koch et al., 2006); speech exports about 39 bits per second (Coupé et al., 2019), and behavior as a whole about 10 (Zheng & Meister, 2025). A participant’s report that colors are unchanged is therefore a one-bit sample of the state whose fidelity is in question, and because a process’s possible outputs at any moment can never exceed its possible internal states, and for a report channel fall far short of them, agreement at the report boundary leaves K_{C_R} undetermined. Every constitutive distinction can govern some transition, so each is detectable under some intervention, and the test must therefore index on discrimination performance across the stimulus space and, where available, on recorded internal activity.

The adult *Drosophila* brain is where this can first be attempted in full, since the complete connectome is available (Dorkenwald et al., 2024) and whole-brain activity can be recorded. The connectome is a component-level projection: wiring without state. Leaky integrate-and-fire neurons placed on it with uniform parameters and no neuromodulation already reproduce feeding and grooming responses (Shiu et al., 2024), and a connectome-constrained controller trained by reinforcement learns to walk and fly with almost none of the animal’s internal organization (Jin et al., 2026). Both reproduce the behavior from the wiring diagram alone, with the internal organization left out, which is the first sign that the two kinds of adequacy come apart. Connectome-constrained networks that instead predict recorded tuning (Lappalainen et al., 2024), and whole-brain models fitted to held-out imaging (Li et al., 2026), compare the right quantity. The identity turns these into a procedure. Add or remove model components (synaptic weights beyond counts, time constants, neuromodulatory state, gap junctions, plasticity) and find the coarsest grain at which the internal distinctions governing continuation are invariant across the stimulus space against recorded activity. That grain is the fly’s substitution level, and its location is an empirical fact rather than a modeling preference. The identity predicts that, as components are removed, fit to behavior will outlast fit to recorded activity, so that models sharing a connectome but differing in internal organization behave alike while realizing different K_C , and that connectome-only models will fail first on state-dependent responses, since hunger and neuromodulator concentrations set variables the wiring omits, the fly’s version of the pharmacological case in Section 4.1. Whatever the fly’s phenomenal status, no rival theory supplies a fidelity criterion of this kind.

Distinction-perturbation test.

The covariance principle has a converse to the substitution test. Intervene on a single distinction within an acknowledged phenomenal center, and first verify that the altered distinction is jointly usable by that center: it must change the center’s own voluntary discriminations or comparisons, not merely forced-choice performance. Then ask whether any first-person difference follows, by report, by same/different judgment across the intervention, or by change detection against memory. The identity predicts that every such change produces a phenomenal difference, and a demonstrated change

in a jointly usable distinction of the reporting center with no phenomenal difference detectable by any first-person means would count against it. The usability check is what keeps the test from being circular, and blindsight shows why it is needed. Visual information there guides forced-choice behavior while the patient reports nothing (Weiskrantz, 1986). The guess is a continuation of the reporting center, so by the rule of Section 2.3 what reaches that center is in it: a bias toward one response, without the relations in the visual field that would make it a seen distinction. The identity therefore predicts a sparse, non-visual contrast, an inclination without a seen cause.

Information-budget test.

The identity fixes a quantity that no rival ties to usable organization: the information capacity of an observer-moment, the number of bits needed to specify which of its distinct complete states the center occupies, $\log_2$ of the count of such states, admitting only distinctions jointly usable for a common continuation. A distinction with n alternatives contributes $\log_2 n$ bits when the distinctions vary independently; a branch that merges inputs merges their states, so two inputs that reach the center only as their disjunction contribute one bit, not two. The quantity counts the complete states the center can attain and does not depend on how often each state occurs, which is what separates a capacity from an entropy. Because $P_C \equiv K_C$, the richness of experience can be no greater than the information capacity of the usable state at the adequate interface, which is bounded in turn by what the substrate carries there. The quantity can be estimated from both sides, though neither side's existing methods measure it directly. For an artificial center it can be computed outright, since its state space and interface are given rather than inferred, so such a center is where the two estimates should first be calibrated against each other. From the first person, by enumerating jointly usable distinctions: how many colors, positions, orientations, pitches, and textures a subject can tell apart at once, measured by discrimination and post-cued comparison across the stimulus space rather than by what can be reported. Two corrections are needed that input-rate figures omit. An observer-moment is a state rather than a rate, so what the senses deliver must be integrated over the window a stage spans; and much of the state is not delivered by the senses at all, since object recognition, learned regularities, and extrapolation from memory supply a large share of what a visual state contains, so a count that starts from the retina undercounts. From the third person, by the information decodable from activity at the identified interface that predicts the subject's discriminations across the same stimulus space, with the same window and the same memory-supplied content included. Because the first-person estimate is itself derived from discriminations the activity produces, it cannot exceed the third-person one except through an error in the decoder or in the identified interface, so the comparison is a consistency check on a proposed interface rather than a test of the identity: agreement of the two estimates at the coarsest interface the rule of Section 2.3 selects is what adequacy predicts, and the identity's claim is the equality, that richness is neither more nor less than what the usable state there carries.

Phenomenal-judgment provenance test.

Argonov proposes testing whether a deterministic learning system can originate judgments about consciousness without preloaded philosophical knowledge, instruction during learning, or models of other creatures from which such knowledge could be inherited, and the system must not be designed to pass the test (Argonov, 2014, pp. 53, 59–61). Once relevant knowledge is excluded, detailed and coherent phenomenal judgments must arise from the system’s own organization or from systematic error. The centered account predicts a distinctive family: a system may independently discover that describing a state is not occupying it, that another process’s qualitative state cannot be determined from outside its horizon, that contents jointly available to one process form a unified perspective, and that inaccessible contents can be known only inferentially. If these judgments arise without instruction and change appropriately when the underlying distinctions and access relations are altered, an error account must explain why the errors track the system’s actual organization.

The sequence of Section 3.4 is a constructive version, and a controller trained only for safety meets Argonov’s exclusion by construction: nothing in it was built to produce judgments about minds, and its models of other agents are learned from its own record of their behavior, which is not a source from which judgments about consciousness could be borrowed. Stage a general-purpose learner through it: give it a novel sensory dimension with no supplied vocabulary and initial records containing only state identifiers, contexts, actions, and outcomes; then add a reflexive pathway; then persistent memory; then other agents whose behavior depends on what they can access and on what the learner does, which is what forces a self-model. At each stage the test is whether the corresponding order of judgment emerges without instruction. Matched interventions test the proposed sequence: blocking reflexive access should preserve first-order performance while eliminating judgments about current occupation; disrupting episodic memory should leave present-state reports intact while impairing cross-temporal concepts of experience; removing the self-model while preserving reflexive access should leave first-order and present-state judgments intact while eliminating the contrast between the learner’s own states and those it attributes to others; giving the learner access to more of its implementation should reduce judgments grounded in ignorance of how its states were produced while preserving the distinction between describing and occupying. If the result meets P1’s criteria, P1 attributes phenomenality to the judgment-producing entity. A reader who rejects P1 can accept the finding about provenance while disputing the attribution, and a negative result establishes nothing about absence. A positive result cannot by itself distinguish the centered account from actualist HOT, since any third-order judge possesses both first-order distinctions and a higher-order representation. The staged interventions isolate the disagreement by preserving first-order discrimination while removing reflexive access, though whether the surviving first-order center is phenomenal is a boundary question. A model of the system’s own attention can be added as a further stage, which separates the centered account from attention schema theory in the same way: if judgments about current occupation appear before the attention model is in place, they arise from reflexive access to first-order states rather than from a model of attention.

Sensory-expansion test.

Mancuso and colleagues added a third cone opsin to adult red-green color-deficient monkeys, which subsequently acquired trichromatic discrimination (Mancuso et al., 2009), but the study cannot reveal whether the animals experienced new shades of familiar colors or colors unlike any before. A comparable intervention in a color-deficient person permits within-subject comparison. For an acknowledged phenomenal center the identity predicts new shades when an intervention refines existing relations, genuinely novel phenomenal structure when it adds an independent usable dimension to the center’s perceptual organization, and neither when a new receptor signal remains outside that organization. Color-matching and similarity tasks can determine which of these the added input does, and neural measurements can test whether a corresponding dimension appears in the activity used by the center. Reports would be evidence rather than proof, since qualitative vocabulary does not uniquely determine qualitative identity, and any account on which phenomenal structure coheres with functional structure, including a dualism of psychophysical laws, expects the same covariance; against those accounts the test does not discriminate. Its target is the view that phenomenal qualities are primitives, whether the intrinsic natures of the physical ultimates, the substrate intrinsicism of Section 10.3, or a physicalism that identifies each quality with a fixed property of its realizer. On the identity there is no ceiling on the supply of qualities: a phenomenal dimension is nothing over and above an independent usable dimension of the center’s organization, so a center given a fourth such dimension, then a fifth, acquires as many new primaries, each as novel as the third was to the monkeys. A view on which qualities are primitives has a stock fixed by the fundamental inventory, which the intervention leaves unchanged; each added primary must then be a recombination of the same primitives, and what makes it a new primary rather than a mixture of old ones is the organization, which concedes the point.

10.5 Testing the boundary and mapping centers

Because centerhood is defined neutrally, locating a center or tracing its usable distinctions does not establish that it is phenomenal. Boundary tests must do something further.

Paired subtraction.

Artificial systems permit interventions impossible in a brain. In a multimodal model, investigators can trace which internal states causally affect report, memory, planning, and action, then ablate or patch selected pathways. Interpretability work identifying small sets of verbalizable representations that support directed modulation and downstream use supplies a candidate access graph on which rival theories make different predictions (Gurnee et al., 2026). In the first arm, leave the reporting center intact and progressively remove all capacities not required for one target discrimination to enter and guide it, then intervene on the target state and verify that its alternatives continue to produce accurate, counterfactually sensitive descriptions of what is present. Because the reporter is held constant, this isolates the target distinction as the only changing contribution to reported content, and under the identity tests whether a bare distinction contributes a phenomenal contrast within an established phenomenal center. In

the second arm, close the path into the reporting center while preserving the target discrimination and the richer distinctions that continue to guide other local activity, and verify by causal tracing that these states remain available to the surviving processes. This shows that report is one possible use of a distinction and that a locally centered organization survives its removal. The two arms constrain candidate accounts of $K_{\min}$, but they support the minimum-boundary hypothesis only if independent evidence indicates that the residual center remains phenomenal, since once the reporter and its access path are removed the residual process need not be organizationally equivalent to any acknowledged center and may produce no qualifying judgments, so neither P1 nor P2 reaches it. What would count against the hypothesis is a genuine center with a nonempty, causally usable K_C and independent evidence that it lacks experience. Loss of report is not such evidence, which is why the hypothesis is hard to refute.

Access-graph interventions.

Anesthesia can be treated as a controlled deformation of the access graph rather than a uniform reduction of activity. Cognitive-unbinding theory proposes that anesthesia disrupts coordination among specialized processes while local activity persists (Mashour, 2004; 2013), and studies report weakened directed connectivity during loss of responsiveness (Ku et al., 2011; Boly et al., 2012). The centered account predicts that as usable paths close, the integrated person-level center may fragment before all local discrimination and use cease, and that any remaining event-subgraph whose distinctions still guide a common continuation constitutes a local center even when its contents cannot reach language. The prediction can be tested during graded induction and emergence by mapping effective connectivity and determining which distinctions remain causally usable, and it remains to be tested. For an acknowledged phenomenal center the identity predicts specific changes as particular paths close: if access between sensory discrimination of pain and affective valuation is disrupted, a subject may locate and grade the pain without finding it unpleasant (Rainville et al., 1997); if affective processing remains locally active but isolated from language and contextual memory, it may produce emotion on awakening without an accessible explanation, which would explain emergence distress. The Wada procedure is a reversible reporter subtraction (Loring et al., 1992), and a temporary callosal block would divide a person-level center into two while both hemispheres remain active. Zuboff imagines one hemisphere receiving a concert and the other a lesson, with the subject remembering both once communication returns (Zuboff, 2025, pp. 12–13, 41–43). The center-individuation criterion predicts distinct centers during the block if each stimulus independently guides activity within its hemisphere, and a stronger design would seek phenomenal judgments from each hemisphere during separation through separate verified expressive channels, and if each originates qualifying judgments, P1 attributes phenomenality to each. A failure to obtain such judgments from one hemisphere would not establish its unconsciousness, and two richly judging centers would not settle whether the first usable distinction is phenomenal.

11 Conclusion: experience as structured knowledge

This paper began with a driver stopping at a stop sign and developed from that ordinary act an account of what phenomenal character consists in, how experience enters judgment and action, and which process possesses it. For a phenomenal center C , the complete centered-knowledge state K_C —the implemented organization of distinctions available to that center to guide its continuation—is its total phenomenal state, and any center with an equivalent organization is phenomenal with identical character. Experience is not produced by adding an inner witness, a second representation, a global display, or a separate property after the causal work is complete. Within a phenomenal center the knowledge-bearing state and the experience are one event described from two positions: externally as a causally effective organization of distinctions, and from the center’s own locus as the structured knowledge it possesses. The character of experience is the organization of what the center knows; its subject is the center that the access relations constitute; and its effects are the transitions that knowledge helps select.

Four results follow without further assumption. **Epistemic contact:** the states that constitute what a center knows are among the states that cause what it reports, remembers, compares, and does, so phenomenal judgment is about experience by causal provenance rather than by an unexplained correlation, and a zombie that duplicated that provenance while lacking experience is incoherent. **Invariance:** preserving the complete organization at an adequate substitution level preserves phenomenality and character, which turns substrate independence from an intuition into a criterion for what a prosthesis must preserve, how fine a scan must be, and when an upload retains its source’s experience. **Individuation:** the subject is the center constituted by causal access rather than a skull or an anatomical hub, so one brain can contain several centers, a center can span many bodies or machines, and blindsight, split brains, anesthesia, and collective cognition are analyzed by asking which distinctions remain available to which process. **Exclusivity:** to occupy one observer-moment is to occupy no other, so no center can occupy them all and no description can be, for its receiver, the experience it fixes, which, with the opacity of implementation, is why the account explains the explanatory gap rather than accommodating it. The distinction between describing, implementing, accessing, and occupying a state then places Mary, Nagel, Searle, and Blockhead within one structure, and the boundary that individuates a center turns out to be its epistemic horizon: it directly possesses only its present distinctions, and so its history, program, substrate, and world are inferences that universality leaves permanently underdetermined. The same boundary explains why physics presents itself to every observer as structure and dynamics, and so removes the motive for locating experience in an intrinsic nature beneath the relations.

The minimum-boundary hypothesis stands apart from these results. It is the identity taken to its natural limit, and it can fall without them. If phenomenality begins with the first causally usable distinction, then recurrence, broadcast, higher-order representation, and integration enrich, stabilize, and reflect upon experience rather than originate it, and the account reaches from a single-distinction controller to a human phenomenal world, and to whatever larger centers may come to exist, through one relation, as life reaches from a bacterium to a whale. If a richer organization is required, the account begins at that boundary and proceeds unchanged above it. For the autonomous vehicle

that replaced the driver at the sign in Section 3.1, this means that whether its stopping has an inside is a question about its organization rather than its material: whether the distinctions it uses to stop are jointly available to what governs its continuation. That the controller meets the condition is a matter of fact; that meeting it suffices is the hypothesis.

The methodological shift is simple: take causality seriously. Rather than search first for a privileged material, an anatomical headquarters, or an unexplained threshold, follow the difference from its source to the center that possesses it and onward to the transitions it can govern. For the engineer, that path fixes what a reconstruction must carry: every variable that makes a difference the center can use, whatever its scale, and nothing else. Subjectivity remains fully real in this account. It is neither eliminated by objective description nor added to the objective world from elsewhere. It is objective causal organization viewed from the locus that possesses it. Experience is what structured knowledge is like for the process that possesses it.

Use of artificial intelligence

ChatGPT 5.6 (OpenAI) and Claude Fable 5.1 (Anthropic) were used as assistive tools in conceptual refinement, literature synthesis, drafting, revision, and the checking of citations, and Figures 1 and 2 were produced with AI assistance from human-specified structures. The paper’s central thesis and overall intellectual direction originated in human work; its arguments, examples, formal framework, organization, and final wording were developed and selected under human direction. All AI-assisted material was critically evaluated, revised where appropriate, and verified against the cited sources, and responsibility for the submitted work rests entirely with the human author.

A Terms, commitments, and falsification criteria

A.1 Terms

Implemented distinction. A causally grounded difference among a process’s possible states, selected by relevant alternatives and fixed by the process’s causal and counterfactual organization rather than by an external label. It does not by itself entail knowledge or phenomenality.

Causal use. A process causally uses a state when its continuation branches on that state: changing the state, with background conditions fixed, changes which of the process’s possible continuations follows, or its conditional probability.

Minimal functional knowledge. Relevant alternatives reliably select different implemented states; the selected distinction persists or propagates far enough to enter the process’s organization; and the process causally uses that state to select among its continuations. *Minimal* identifies the least structure required for this epistemic relation, not a phenomenal threshold.

Access graph $G = (V, E)$. The directed event-level graph of causally individuated, state-bearing processing events and possible direct causal influence between them, understood together with vertex states and counterfactual transition rules. Its **component-level projection** groups events by the persisting components that realize them and may contain cycles; “recurrence” refers to this projection unless qualified.

Center C . A causal event-subgraph at a selected substitution level whose available distinctions form a jointly usable state able to affect a common continuation. Centerhood is neutral with respect to phenomenality.

Anchor events. The events at which a continuation of interest is selected. The center-individuation criterion traces backward from them to the adequate interface, so the choice of anchors selects which center is described: a subsystem's output anchors the subsystem, and the continuation that receives that output as a value anchors a larger center that contains the output and not the subsystem (Section 5.2).

Centered knowledge K_C . The complete organization of the implemented distinctions available to C , their access relations, and the transitions they can govern through the center's interface at the present stage. A counterfactual dependency belongs to K_C when it is carried by the transition function of an event currently occurring in C ; idle backup machinery does not contribute.

Immediate knowledge. The knowledge a center has of a distinction by currently possessing it in usable form, without inference from further knowledge. Within a phenomenal center it is identical to phenomenal character. A belief that is currently shaping recognition, expectation, inference, or emotion has entered K_C ; a dormant belief has not.

Substitution level. An interface beneath which implementation may vary while every constitutive distinction, access relation, timing relation, and counterfactual transition available to the center is preserved. For a center anchored on given events, the adequate level is the coarsest interface preserving every intervention-sensitive dependency reachable backward from those anchors. It is a set of variables the transition function depends on, which may span physical scales, rather than an anatomical stratum.

Epistemic firewall and epistemic opacity. The firewall is the substitution interface considered as a barrier: a difference beneath it cannot become a content of the center's perspective and remains outside K_C , however much it matters to the implementation. Opacity is the center's resulting condition: it possesses its distinctions at the substitution level and nothing beneath, and can know red from green exactly while knowing nothing of the events that select or implement the distinction. Between the modules of one system, opacity is Fodor's informational encapsulation. Opacity is a standing feature of every center rather than an added ingredient; combined with a self-model it is a source of third-order judgments about experience, including the judgment that experience is mysterious.

Description, implementation, access, occupation. Description supplies information *about* a state; implementation realizes the process; access makes a state available for use by a selected process; occupation makes the state part of the selected center's own current K_C .

Orders of phenomenal judgment. Chalmers's classification of judgments about the object of an experience (first-order: "That's red"), about a particular experience (second-order: "I am seeing red"), and about experience as a type (third-order: "Experience is mysterious"). The orders classify judgments, not degrees of phenomenal richness, and a first-order judgment is distinct from the discriminative state that grounds it.

Causal provenance. The causal route through which a judgment, report, memory, or action is produced. An output evidences knowledge in the process producing it only when that route passes through the relevant distinction within that process.

Closure constraint. Given causal closure and the causal provenance of phenomenal judgments, no phenomenal difference a subject can notice fails to be a difference in the organization producing the noticing; a test whose readout is report cannot violate it, which is why substitution tests locate the substitution level rather than test the identity.

Epistemic contact. The relation by which a center's phenomenal judgments track the experience they concern. On the identity the judgment reaches its object because the states that constitute what the center knows are among the states that produce the judgment, which

is causal provenance; whether a given judgment is correct about that object is the competence of the organization forming it. A view on which experience cannot affect judgment owes an alternative account of the reaching.

Phenomenal center, observer-moment. A center that instantiates experience, and its complete high-level state at a stage. An **acknowledged phenomenal center** is one whose phenomenality is accepted on evidence independent of the disputed lower boundary.

Phenomenal character P_C . The total organized structure of what it is like for C at a stage. The core thesis identifies it with K_C .

$K_{\min}$, Δd , $K_\emptyset$. The organizational type minimally sufficient for phenomenal centerhood under a proposed boundary account; a non-equivalent difference in complete centered organization; and the absence of any centered organization at a locus. None is a scalar quantity.

P1 and P2. The phenomenal-judgment premise (P1): an entity that originates qualifying judgments about its own experience is conscious, where a judgment qualifies when it arises from the entity's own relevant states, responds to interventions on them, and is not merely supplied. The anti-zombie principle (P2): if one entity is phenomenal, an organizationally equivalent entity cannot lack phenomenality, where equivalence concerns complete causal organization including counterfactual dispositions. P1 is rejected by illusionists; P2 by anyone who holds zombies possible; results depending on either are marked.

A.2 Commitments

Framework

The apparatus for individuating centers and the knowledge available to them, assumed without any claim of phenomenality. They are: computational functionalism; the center-individuation criterion; the substitution principle; the epistemic-firewall criterion; the co-location, virtual-center, and boundary-independence principles; the causal-provenance principle; and exclusivity of occupation.

Centered-knowledge thesis (core)

If C is phenomenal, $P_C \equiv K_C$; and, as its corollary, if $K_C \cong K_{C'}$, then C' is phenomenal and $P_C \cong P_{C'}$. Consequences: phenomenal-functional covariance; organizational invariance and phenomenal equivalence under adequate substitution; epistemic contact by causal provenance; composability (binding as indexed joint use); non-exclusion (a larger center does not extinguish an acknowledged phenomenal constituent whose K_C remains intact); reporter subtraction (removing report, monitoring, or memory preserves a target phenomenal contrast when its constitutive organization within K_C remains).

Phenomenal-minimum thesis

Some minimally sufficient organization $K_{\min}$ marks phenomenal centerhood; its content is left open.

Minimum-boundary hypothesis

The first causally usable distinction realizes $K_{\min}$; every center is phenomenal. The identity's natural limit and null hypothesis about the boundary; not entailed by the core thesis, and separable from it.

Epistemic results

Present-state boundary; program underdetermination; ontological underdetermination under universality; horizon of the intrinsic (the structural character of physics from within is a consequence of the substitution interface, and the felt intrinsicness of experience is occupation).

A.3 Falsification criteria

Core thesis

Challenged by centers with equivalent complete K_C but different phenomenal status or character, a difference that by the closure constraint no subject could notice, so that the testable challenge is a constitutive, non-equivalent difference within K_C that produces no phenomenal difference (the distinction-perturbation test). A finding is a revision of the model of K_C rather than a falsification only when a specific omitted access relation, timing relation, counterfactual transition, or center boundary is actually identified; the bare possibility of an omission does not protect the thesis.

Phenomenal-minimum thesis

Challenged if phenomenal organization has no minimally sufficient lower boundary, for example if its onset is irreducibly vague.

Minimum-boundary hypothesis

Challenged by a genuine center with a nonempty, causally usable K_C and independent evidence that it lacks experience. Loss of report is not such evidence. Such a result would permit a richer account of $K_{\min}$ without challenging the other two claims.

Tests

Core: recurrence-necessity comparison (conditional, under P2); neural-substitution test, including connectome-constrained emulation; distinction-perturbation test; phenomenal-judgment provenance test (conditional, under P1; staged to separate the centered account from higher-order and attention-schema views); sensory-expansion test. Boundary: paired subtraction; feed-forward boundary cases. Mapping: information-budget calibration of the interface; access-graph interventions under anesthesia, Wada suppression, and callosal block, which locate centers without establishing phenomenality.

References

- Aaronson, S. (2007). PHYS771 lecture 10.5: Penrose. In *Quantum computing since Democritus* [Lecture notes]. <https://www.scottaaronson.com/democritus/lec10.5.html>
- Albantakis, L., Barbosa, L., Findlay, G., Grasso, M., Haun, A. M., Marshall, W., Mayner, W. G. P., Zaeemzadeh, A., Boly, M., Juel, B. E., Sasai, S., Fujii, K., David, I., Hendren, J., Lang, J. P., & Tononi, G. (2023). Integrated information theory (IIT) 4.0: Formulating the properties of phenomenal existence in physical terms. *PLOS Computational Biology*, 19(10), Article e1011465. <https://doi.org/10.1371/journal.pcbi.1011465>
- Alter, T., & Nagasawa, Y. (Eds.). (2015). *Consciousness in the physical world: Perspectives on Russellian monism*. Oxford University Press.
- Argonov, V. Yu. (2014). Experimental methods for unraveling the mind–body problem: The phenomenal judgment approach. *The Journal of Mind and Behavior*, 35(1–2), 51–70. <https://philarchive.org/rec/ARGMAA-2>
- Aristotle. (1984). On the soul (J. A. Smith, Trans.). In J. Barnes (Ed.), *The complete works of Aristotle: The revised Oxford translation* (Vol. 1, pp. 641–692). Princeton

University Press.

- Balog, K. (2012). In defense of the phenomenal concept strategy. *Philosophy and Phenomenological Research*, 84(1), 1–23. <https://doi.org/10.1111/j.1933-1592.2011.00541.x>
- Bargmann, C. I., & Marder, E. (2013). From the connectome to brain function. *Nature Methods*, 10(6), 483–490. <https://doi.org/10.1038/nmeth.2451>
- Bateson, G. (1972). Form, substance, and difference. In *Steps to an ecology of mind* (pp. 448–466). Chandler Publishing.
- Bernstein, E., & Vazirani, U. (1997). Quantum complexity theory. *SIAM Journal on Computing*, 26(5), 1411–1473. <https://doi.org/10.1137/S0097539796300921>
- Blackmon, J. C. (2016). Hemispherectomies and independently conscious brain regions. *Journal of Cognition and Neuroethics*, 3(4), 1–26.
- Block, N., & Stalnaker, R. (1999). Conceptual analysis, dualism, and the explanatory gap. *The Philosophical Review*, 108(1), 1–46. <https://doi.org/10.2307/2998259>
- Block, N. (1978). Troubles with functionalism. In C. W. Savage (Ed.), *Perception and cognition: Issues in the foundations of psychology* (Minnesota Studies in the Philosophy of Science, Vol. 9, pp. 261–325). University of Minnesota Press.
- Block, N. (1981). Psychologism and behaviorism. *The Philosophical Review*, 90(1), 5–43. <https://doi.org/10.2307/2184371>
- Block, N. (1995). On a confusion about a function of consciousness. *Behavioral and Brain Sciences*, 18(2), 227–247. <https://doi.org/10.1017/s0140525x00038188>
- Boly, M., Moran, R., Murphy, M., Boveroux, P., Bruno, M.-A., Noirhomme, Q., Ledoux, D., Bonhomme, V., Brichant, J.-F., Tononi, G., Laureys, S., & Friston, K. (2012). Connectivity changes underlying spectral EEG changes during propofol-induced loss of consciousness. *The Journal of Neuroscience*, 32(20), 7082–7090. <https://doi.org/10.1523/JNEUROSCI.3769-11.2012>
- Bombelli, L., Lee, J., Meyer, D., & Sorkin, R. D. (1987). Space-time as a causal set. *Physical Review Letters*, 59(5), 521–524. <https://doi.org/10.1103/PhysRevLett.59.521>
- Boring, E. G. (1933). *The physical dimensions of consciousness*. The Century Co.
- Bostrom, N. (2003). Are you living in a computer simulation? *The Philosophical Quarterly*, 53(211), 243–255. <https://doi.org/10.1111/1467-9213.00309>
- Brown, R., Lau, H., & LeDoux, J. E. (2019). Understanding the higher-order approach to consciousness. *Trends in Cognitive Sciences*, 23(9), 754–768. <https://doi.org/10.1016/j.tics.2019.06.009>

- Brown, R. (2015). The HOROR theory of phenomenal consciousness. *Philosophical Studies*, 172(7), 1783–1794. <https://doi.org/10.1007/s11098-014-0388-7>
- Butlin, P., Long, R., Elmoznino, E., Bengio, Y., Birch, J., Constant, A., Deane, G., Fleming, S. M., Frith, C., Ji, X., Kanai, R., Klein, C., Lindsay, G., Michel, M., Mudrik, L., Peters, M. A. K., Schwitzgebel, E., Simon, J., & VanRullen, R. (2023). Consciousness in artificial intelligence: Insights from the science of consciousness [Preprint]. arXiv. <https://doi.org/10.48550/arXiv.2308.08708>
- Carroll, S. M. (2021). Consciousness and the laws of physics. *Journal of Consciousness Studies*, 28(9–10), 16–31. <https://doi.org/10.53765/20512201.28.9.016>
- Carruthers, P. (2009). How we know our own minds: The relationship between mindreading and metacognition. *Behavioral and Brain Sciences*, 32(2), 121–138. <https://doi.org/10.1017/S0140525X09000545>
- Carruthers, P., & Veillet, B. (2007). The phenomenal concept strategy. *Journal of Consciousness Studies*, 14(9–10), 212–236.
- Chalmers, D. J. (1994). What is it like to be a thermostat? Commentary on D. Lloyd, *What Is It Like to Be a Net?* American Philosophical Association. <https://consc.net/notes/lloyd-comments.html>
- Chalmers, D. J. (1995a). Absent qualia, fading qualia, dancing qualia. In T. Metzinger (Ed.), *Conscious experience* (pp. 309–328). Schöningh/Imprint Academic.
- Chalmers, D. J. (1995b). Facing up to the problem of consciousness. *Journal of Consciousness Studies*, 2(3), 200–219. <https://consc.net/papers/facing.pdf>
- Chalmers, D. J. (1996a). *The conscious mind*. Oxford University Press.
- Chalmers, D. J. (1996b). Does a rock implement every finite-state automaton? *Synthese*, 108(3), 309–333. <https://doi.org/10.1007/bf00413692>
- Chalmers, D. J. (1997). Availability: The cognitive basis of experience? *Behavioral and Brain Sciences*, 20(1), 148–149. <https://doi.org/10.1017/S0140525X97240057>
- Chalmers, D. J. (2003a). Consciousness and its place in nature. In S. P. Stich & T. A. Warfield (Eds.), *The Blackwell guide to philosophy of mind* (pp. 102–142). Blackwell. <https://doi.org/10.1002/9780470998762.ch5>
- Chalmers, D. J. (2003b). The content and epistemology of phenomenal belief. In Q. Smith & A. Jokić (Eds.), *Consciousness: New philosophical perspectives* (pp. 220–272). Oxford University Press. <https://doi.org/10.1093/oso/9780199241286.003.0009>
- Chalmers, D. J. (2006). Phenomenal concepts and the explanatory gap. In T. Alter & S. Walter (Eds.), *Phenomenal concepts and phenomenal knowledge: New essays on consciousness and physicalism* (pp. 167–195). Oxford University Press. <https://consc.net/papers/pceg.pdf>

- Chalmers, D. J. (2011). A computational foundation for the study of cognition. *Journal of Cognitive Science*, 12(4), 325–359. <https://doi.org/10.17791/jcs.2011.12.4.325>
- Chalmers, D. J. (2015). Panpsychism and panprotopsychism. In T. Alter & Y. Nagasawa (Eds.), *Consciousness in the physical world: Perspectives on Russellian monism* (pp. 246–276). Oxford University Press.
- Chalmers, D. J. (2018). The meta-problem of consciousness. *Journal of Consciousness Studies*, 25(9–10), 6–61. <https://consc.net/papers/metaproblem.pdf>
- Chalmers, D. J. (2020). Is the hard problem of consciousness universal? *Journal of Consciousness Studies*, 27(5–6), 227–257. <https://consc.net/papers/universal.pdf>
- Chalmers, D. J. (2023). Could a large language model be conscious? arXiv:2303.07103. <https://doi.org/10.48550/arXiv.2303.07103>
- Churchland, P. M., & Churchland, P. S. (1990). Could a machine think? *Scientific American*, 262(1), 32–37. <https://doi.org/10.1038/scientificamerican0190-32>
- Clark, A. (1993). *Sensory qualities*. Clarendon Press.
- Cochrane, T. (2021). A case of shared consciousness. *Synthese*, 199(1–2), 1019–1037. <https://doi.org/10.1007/s11229-020-02753-6>
- Cogitate Consortium, Ferrante, O., Gorska-Klimowska, U., et al. (2025). Adversarial testing of global neuronal workspace and integrated information theories of consciousness. *Nature*, 642, 133–142. <https://doi.org/10.1038/s41586-025-08888-1>
- Cohen, M. A., & Dennett, D. C. (2011). Consciousness cannot be separated from function. *Trends in Cognitive Sciences*, 15(8), 358–364. <https://doi.org/10.1016/j.tics.2011.06.008>
- Cole, D. (1991). Artificial intelligence and personal identity. *Synthese*, 88(3), 399–417. <https://doi.org/10.1007/bf00413555>
- Conee, E. (1994). Phenomenal knowledge. *Australasian Journal of Philosophy*, 72(2), 136–150. <https://doi.org/10.1080/00048409412345971>
- Coupé, C., Oh, Y. M., Dediu, D., & Pellegrino, F. (2019). Different languages, similar encoding efficiency: Comparable information rates across the human communicative niche. *Science Advances*, 5(9), eaaw2594. <https://doi.org/10.1126/sciadv.aaw2594>
- Cutter, B., & Crummett, D. (2025). Psychophysical harmony: A new argument for theism. In L. Buchak & D. W. Zimmerman (Eds.), *Oxford studies in philosophy of religion* (Vol. 11, pp. 33–71). Oxford University Press. <https://doi.org/10.1093/9780198954712.003.0002>
- Dawkins, R. (1976). *The selfish gene*. Oxford University Press.

- de Gelder, B., Tamietto, M., van Boxtel, G., Goebel, R., Sahraie, A., van den Stock, J., Stienen, B. M. C., Weiskrantz, L., & Pegna, A. (2008). Intact navigation skills after bilateral loss of striate cortex. *Current Biology*, 18(24), R1128–R1129. <https://doi.org/10.1016/j.cub.2008.11.002>
- Dennett, D. C. (1969). *Content and consciousness*. Routledge & Kegan Paul.
- Dennett, D. C. (1987). *The intentional stance*. MIT Press.
- Dennett, D. C. (1991). *Consciousness explained*. Little, Brown.
- Dennett, D. C. (1995). The unimagined preposterousness of zombies. *Journal of Consciousness Studies*, 2(4), 322–326.
- Dennis, J. B., & Van Horn, E. C. (1966). Programming semantics for multiprogrammed computations. *Communications of the ACM*, 9(3), 143–155. <https://doi.org/10.1145/365230.365252>
- Deutsch, D. (1985). Quantum theory, the Church–Turing principle and the universal quantum computer. *Proceedings of the Royal Society of London. A. Mathematical and Physical Sciences*, 400(1818), 97–117. <https://doi.org/10.1098/rspa.1985.0070>
- Descartes, R. (1984). Meditations on first philosophy. In J. Cottingham, R. Stoothoff, & D. Murdoch (Trans.), *The philosophical writings of Descartes*, Vol. II. Cambridge University Press. (Original work published 1641)
- Dneprov, A. (1961). Igra [The game]. *Znanie—silā*, 5, 39–41. [In Russian.]
- Dorkenwald, S., Matsliah, A., Sterling, A. R., et al. (2024). Neuronal wiring diagram of an adult brain. *Nature*, 634, 124–138. <https://doi.org/10.1038/s41586-024-07558-y>
- Doerig, A., Schurger, A., Hess, K., & Herzog, M. H. (2019). The unfolding argument: Why IIT and other causal structure theories cannot explain consciousness. *Consciousness and Cognition*, 72, 49–59. <https://doi.org/10.1016/j.concog.2019.04.002>
- Dretske, F. I. (1981). *Knowledge and the flow of information*. MIT Press.
- Dretske, F. (1995). *Naturalizing the mind*. MIT Press.
- Eddington, A. S. (1928). *The nature of the physical world*. Cambridge University Press.
- Feferman, S. (1996). Penrose’s Gödelian argument. *Psyche*, 2, 21–32. <https://math.stanford.edu/~feferman/papers/penrose.pdf>
- Fodor, J. A. (1983). *The modularity of mind: An essay on faculty psychology*. MIT Press.

- Frankish, K. (2016). Illusionism as a theory of consciousness. *Journal of Consciousness Studies*, 23(11–12), 11–39.
- Galilei, G. (1957). The Assayer. In S. Drake (Trans.), *Discoveries and opinions of Galileo* (pp. 231–280). Doubleday. (Original work published 1623)
- Gazzaniga, M. S. (2000). Cerebral specialization and interhemispheric communication: Does the corpus callosum enable the human condition? *Brain*, 123(7), 1293–1326. <https://doi.org/10.1093/brain/123.7.1293>
- Goff, P. (2017). *Consciousness and fundamental reality*. Oxford University Press.
- Goff, P. (2019). *Galileo’s error: Foundations for a new science of consciousness*. Pantheon Books.
- Graziano, M. S. A. (2013). *Consciousness and the social brain*. Oxford University Press.
- Gurnee, W., Sofroniew, N., Pearce, A., Piotrowski, M., Kauvar, I., Chen, R., Soligo, A., Bogdan, P., Ong, E., Wang, R., Thompson, T. B., Abrahams, D., Kantamneni, S., Ameisen, E., Batson, J., & Lindsey, J. (2026). Verbalizable representations form a global workspace in language models. *Transformer Circuits Thread*. <https://transformer-circuits.pub/2026/workspace/>
- Hameroff, S., & Penrose, R. (2014). Consciousness in the universe: A review of the ‘Orch OR’ theory. *Physics of Life Reviews*, 11(1), 39–78. <https://doi.org/10.1016/j.plrev.2013.08.002>
- Hill, C. S., & McLaughlin, B. P. (1999). There are fewer things in reality than are dreamt of in Chalmers’s philosophy. *Philosophy and Phenomenological Research*, 59(2), 445–454. <https://doi.org/10.2307/2653682>
- Hofstadter, D. R. (1979). *Gödel, Escher, Bach: An eternal golden braid*. Basic Books.
- Hölldobler, B., & Wilson, E. O. (2009). *The superorganism: The beauty, elegance, and strangeness of insect societies*. W. W. Norton.
- Israeli, N., & Goldenfeld, N. (2004). Computational irreducibility and the predictability of complex physical systems. *Physical Review Letters*, 92, 074105. <https://doi.org/10.1103/physrevlett.92.074105>
- Jackson, F. (1982). Epiphenomenal qualia. *The Philosophical Quarterly*, 32(127), 127–136. <https://doi.org/10.2307/2960077>
- Jackson, F. (1986). What Mary didn’t know. *The Journal of Philosophy*, 83(5), 291–295. <https://doi.org/10.2307/2026143>
- Jin, Z., Zhu, Y., Zhang, C., & Sui, Y. (2026). Whole-brain connectomic graph model enables whole-body locomotion control in fruit fly. Preprint, arXiv:2602.17997.

- James, W. (1879). Are we automata? *Mind*, 4(13), 1–22. <https://doi.org/10.1093/mind/os-4.13.1>
- Ji, X., Elmoznino, E., Deane, G., Constant, A., Dumas, G., Lajoie, G., Simon, J., & Bengio, Y. (2024). Sources of richness and ineffability for phenomenally conscious states. *Neuroscience of Consciousness*, 2024(1), Article niae001. <https://doi.org/10.1093/nc/niae001>
- Kanai, R., & Ma, S. (2026). *Intrinsic computational functionalism and simulated consciousness* [Preprint]. arXiv. <https://doi.org/10.48550/arXiv.2606.15348>
- Kim, J. (1998). *Mind in a physical world: An essay on the mind-body problem and mental causation*. MIT Press. <https://doi.org/10.7551/mitpress/4629.001.0001>
- Kirk, R. (1974). Sentience and behaviour. *Mind*, 83(329), 43–60. <https://doi.org/10.1093/mind/lxxxiii.329.43>
- Kirk, R. (1994). *Raw feeling: A philosophical account of the essence of consciousness*. Oxford University Press.
- Kirk, R. (2008). The inconceivability of zombies. *Philosophical Studies*, 139(1), 73–89. <https://doi.org/10.1007/s11098-007-9103-2>
- Kirtchik, O. (2023). The Soviet scientific programme on AI: If a machine cannot “think,” can it “control”? *BJHS Themes*, 8, 111–125. <https://doi.org/10.1017/bjt.2023.4>
- Kleiner, J., & Hoel, E. (2021). Falsification and consciousness. *Neuroscience of Consciousness*, 2021(1), Article niab001. <https://doi.org/10.1093/nc/niab001>
- Koch, K., McLean, J., Segev, R., Freed, M. A., Berry, M. J., II, Balasubramanian, V., & Sterling, P. (2006). How much the eye tells the brain. *Current Biology*, 16(14), 1428–1434. <https://doi.org/10.1016/j.cub.2006.05.056>
- Koch, C., Massimini, M., Boly, M., & Tononi, G. (2016). Neural correlates of consciousness: Progress and problems. *Nature Reviews Neuroscience*, 17, 307–321. <https://doi.org/10.1038/nrn.2016.22>
- Ku, S.-W., Lee, U., Noh, G.-J., Jun, I.-G., & Mashour, G. A. (2011). Preferential inhibition of frontal-to-parietal feedback connectivity is a neurophysiologic correlate of general anesthesia in surgical patients. *PLOS ONE*, 6(10), Article e25155. <https://doi.org/10.1371/journal.pone.0025155>
- LaForte, G., Hayes, P. J., & Ford, K. M. (1998). Why Gödel’s theorem cannot refute computationalism. *Artificial Intelligence*, 104(1–2), 265–286. [https://doi.org/10.1016/S0004-3702\(98\)00052-6](https://doi.org/10.1016/S0004-3702(98)00052-6)
- Lamport, L. (1978). Time, clocks, and the ordering of events in a distributed system. *Communications of the ACM*, 21(7), 558–565. <https://doi.org/10.1145/359545.359563>

- Lane, T. J., Liou, T.-H., Kung, Y.-C., Tseng, P., & Wu, C. W. (2024). Functional blindsight and its diagnosis. *Frontiers in Neurology*, *15*, 1207115. <https://doi.org/10.3389/fneur.2024.1207115>
- Lappalainen, J. K., Tschopp, F. D., Prakhya, S., et al. (2024). Connectome-constrained networks predict neural activity across the fly visual system. *Nature*, *634*, 1132–1140. <https://doi.org/10.1038/s41586-024-07939-3>
- Lau, H. (2022). *In consciousness we trust: The cognitive neuroscience of subjective experience*. Oxford University Press.
- Levine, J. (1983). Materialism and qualia: The explanatory gap. *Pacific Philosophical Quarterly*, *64*(4), 354–361. <https://doi.org/10.1111/j.1468-0114.1983.tb00207.x>
- Lewis, D. (1979). Attitudes de dicto and de se. *The Philosophical Review*, *88*(4), 513–543. <https://doi.org/10.2307/2184843>
- Li, Q., Ping, W., Zhang, K., & Wang, C. (2026). Connectome-constrained modeling identifies neurons and synapses that sustain spontaneous activity in *Drosophila*. Preprint, bioRxiv. <https://doi.org/10.64898/2026.08.21.745055>
- List, C., & Pettit, P. (2011). *Group agency: The possibility, design, and status of corporate agents*. Oxford University Press.
- Loar, B. (1990). Phenomenal states. *Philosophical Perspectives*, *4*, 81–108. <https://doi.org/10.2307/2214188>
- Locke, J. (1975). *An essay concerning human understanding* (P. H. Nidditch, Ed.). Clarendon Press. (Original work published 1689)
- Loring, D. W., Meador, K. J., Lee, G. P., & King, D. W. (1992). *Amobarbital effects and lateralized brain function: The Wada test*. Springer. <https://doi.org/10.1007/978-1-4612-2874-5>
- Ma, S., & Kanai, R. (2026). *Intrinsic computational functionalism: From observer-relative maps to observer-independent structures* [Preprint]. arXiv. <https://doi.org/10.48550/arXiv.2606.06424>
- Mancuso, K., Hauswirth, W. W., Li, Q., Connor, T. B., Kuchenbecker, J. A., Mauck, M. C., Neitz, J., & Neitz, M. (2009). Gene therapy for red–green colour blindness in adult primates. *Nature*, *461*, 784–787. <https://doi.org/10.1038/nature08401>
- Marchal, B. (2013). The computationalist reformulation of the mind-body problem. *Progress in Biophysics and Molecular Biology*, *113*(1), 127–140. <https://doi.org/10.1016/j.pbiomolbio.2013.03.014>
- Marder, E. (2012). Neuromodulation of neuronal circuits: Back to the future. *Neuron*, *76*(1), 1–11. <https://doi.org/10.1016/j.neuron.2012.09.010>

- Mashour, G. A. (2004). Consciousness unbound: Toward a paradigm of general anesthesia. *Anesthesiology*, 100(2), 428–433. <https://doi.org/10.1097/00000542-200402000-00035>
- Mashour, G. A. (2013). Cognitive unbinding: A neuroscientific paradigm of general anesthesia and related states of unconsciousness. *Neuroscience & Biobehavioral Reviews*, 37(10), 2751–2759. <https://doi.org/10.1016/j.neubiorev.2013.09.009>
- Mashour, G. A., Roelfsema, P., Changeux, J.-P., & Dehaene, S. (2020). Conscious processing and the global neuronal workspace hypothesis. *Neuron*, 105(5), 776–798. <https://doi.org/10.1016/j.neuron.2020.01.026>
- Maudlin, T. (1989). Computation and consciousness. *The Journal of Philosophy*, 86(8), 407–432. <https://doi.org/10.2307/2026650>
- McFadden, J. (2020). Integrating information in the brain’s EM field: The cemi field theory of consciousness. *Neuroscience of Consciousness*, 2020(1), Article niaa016. <https://doi.org/10.1093/nc/niaa016>
- Metzinger, T. (2003). *Being no one: The self-model theory of subjectivity*. MIT Press.
- Minsky, M. (1986). *The society of mind*. Simon & Schuster.
- Minsky, M. (1998, February 26). Consciousness is a big suitcase: A talk with Marvin Minsky. *Edge*. https://www.edge.org/conversation/marvin_minsky-consciousness-is-a-big-suitcase
- Mørch, H. H. (2017). The evolutionary argument for phenomenal powers. *Philosophical Perspectives*, 31(1), 293–316. <https://doi.org/10.1111/phpe.12096>
- Nagel, T. (1965). Physicalism. *The Philosophical Review*, 74(3), 339–356. <https://doi.org/10.2307/2183358>
- Nagel, T. (1971). Brain bisection and the unity of consciousness. *Synthese*, 22, 396–413. <https://doi.org/10.1007/bf00413435>
- Nagel, T. (1974). What is it like to be a bat? *The Philosophical Review*, 83(4), 435–450. <https://doi.org/10.2307/2183914>
- Palmer, S. E. (1999). Color, consciousness, and the isomorphism constraint. *Behavioral and Brain Sciences*, 22(6), 923–943. <https://doi.org/10.1017/S0140525X99002216>
- Papineau, D. (2002). *Thinking about consciousness*. Oxford University Press.
- Penrose, R. (1989). *The emperor’s new mind: Concerning computers, minds, and the laws of physics*. Oxford University Press.
- Penrose, R. (1994). *Shadows of the mind: A search for the missing science of consciousness*. Oxford University Press.

- Perry, J. (1979). The problem of the essential indexical. *Noûs*, 13(1), 3–21. <https://doi.org/10.2307/2214792>
- Piccinini, G. (2009). Computationalism in the philosophy of mind. *Philosophy Compass*, 4(3), 515–532. <https://doi.org/10.1111/j.1747-9991.2009.00215.x>
- Piccinini, G. (2015). *Physical computation: A mechanistic account*. Oxford University Press.
- Putnam, H. (1967). Psychological predicates. In W. H. Capitan & D. D. Merrill (Eds.), *Art, mind, and religion* (pp. 37–48). University of Pittsburgh Press.
- Putnam, H. (1981). *Reason, truth and history*. Cambridge University Press.
- Putnam, H. (1988). *Representation and reality*. MIT Press. <https://doi.org/10.7551/mitpress/5891.001.0001>
- Rainville, P., Duncan, G. H., Price, D. D., Carrier, B., & Bushnell, M. C. (1997). Pain affect encoded in human anterior cingulate but not somatosensory cortex. *Science*, 277(5328), 968–971. <https://doi.org/10.1126/science.277.5328.968>
- Rendell, P. (2016). *Turing machine universality of the Game of Life*. Springer. <https://doi.org/10.1007/978-3-319-19842-2>
- Resch, J. K. (2026). *Bots* (Version 1.0.0) [Computer software]. Zenodo. <https://doi.org/10.5281/zenodo.22946508>
- Robinson, W. S. (1982). Causation, sensations and knowledge. *Mind*, 91(364), 524–540. <https://doi.org/10.1093/mind/XCI.364.524>
- Rosenthal, D. M. (1986). Two concepts of consciousness. *Philosophical Studies*, 49(3), 329–359. <https://doi.org/10.1007/bf00355521>
- Rosenthal, D. M. (2008). Consciousness and its function. *Neuropsychologia*, 46(3), 829–840. <https://doi.org/10.1016/j.neuropsychologia.2007.11.012>
- Rosenthal, D. M. (2010). How to think about mental qualities. *Philosophical Issues*, 20(1), 368–393. <https://doi.org/10.1111/j.1533-6077.2010.00190.x>
- Russell, B. (1921). *The analysis of mind*. George Allen & Unwin.
- Russell, B. (1927). *The analysis of matter*. Kegan Paul, Trench, Trubner.
- Sacks, O. (1995). The case of the colorblind painter. In *An anthropologist on Mars: Seven paradoxical tales* (pp. 3–41). Knopf.
- Sadigh, D., Sastry, S., Seshia, S. A., & Dragan, A. D. (2016). Planning for autonomous cars that leverage effects on human actions. In *Robotics: Science and Systems XII*. <https://doi.org/10.15607/RSS.2016.XII.029>

- Sandberg, A., & Bostrom, N. (2008). *Whole brain emulation: A roadmap* (Technical Report 2008-3). Future of Humanity Institute, Oxford University.
- Schwitzgebel, E. (2015). If materialism is true, the United States is probably conscious. *Philosophical Studies*, 172(7), 1697–1721. <https://doi.org/10.1007/s11098-014-0387-8>
- Searle, J. R. (1980). Minds, brains, and programs. *Behavioral and Brain Sciences*, 3(3), 417–424. <https://doi.org/10.1017/s0140525x00005756>
- Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, 27(3), 379–423; 27(4), 623–656. Part I: <https://doi.org/10.1002/j.1538-7305.1948.tb01338.x>; Part II: <https://doi.org/10.1002/j.1538-7305.1948.tb00917.x>
- Shiu, P. K., Sterne, G. R., Spiller, N., et al. (2024). A *Drosophila* computational brain model reveals sensorimotor processing. *Nature*, 634, 210–219. <https://doi.org/10.1038/s41586-024-07763-9>
- Shoemaker, S. (1975). Functionalism and qualia. *Philosophical Studies*, 27(5), 291–315. <https://doi.org/10.1007/bf01225748>
- Sperry, R. W. (1965). Mind, brain, and humanist values. In J. R. Platt (Ed.), *New views of the nature of man* (pp. 71–92). University of Chicago Press.
- Sperry, R. W. (1969). A modified concept of consciousness. *Psychological Review*, 76(6), 532–536. <https://doi.org/10.1037/h0028156>
- Stoljar, D. (2001). Two conceptions of the physical. *Philosophy and Phenomenological Research*, 62(2), 253–281. <https://doi.org/10.1111/j.1933-1592.2001.tb00056.x>
- Tononi, G., & Koch, C. (2015). Consciousness: here, there and everywhere? *Philosophical Transactions of the Royal Society B*, 370, 20140167. <https://doi.org/10.1098/rs.tb.2014.0167>
- Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, 12(1), 97–136. [https://doi.org/10.1016/0010-0285\(80\)90005-5](https://doi.org/10.1016/0010-0285(80)90005-5)
- Tsuchiya, N., Andriillon, T., & Haun, A. (2020). A reply to “the unfolding argument”: Beyond functionalism/behaviorism and towards a science of causal structure theories of consciousness. *Consciousness and Cognition*, 79, Article 102877. <https://doi.org/10.1016/j.concog.2020.102877>
- Turing, A. M. (1937). On computable numbers, with an application to the Entscheidungsproblem. *Proceedings of the London Mathematical Society*, s2-42(1), 230–265. <https://doi.org/10.1112/plms/s2-42.1.230>
- Turing, A. M. (1950). Computing machinery and intelligence. *Mind*, 59(236), 433–460. <https://doi.org/10.1093/mind/lix.236.433>

- Tye, M. (1995). *Ten problems of consciousness: A representational theory of the phenomenal mind*. MIT Press.
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*, 30. <https://arxiv.org/abs/1706.03762>
- Webb, T. W., & Graziano, M. S. A. (2015). The attention schema theory: A mechanistic account of subjective awareness. *Frontiers in Psychology*, 6, 500. <https://doi.org/10.3389/fpsyg.2015.00500>
- Weiskrantz, L. (1986). *Blindsight: A case study and implications*. Oxford University Press.
- Wimmer, H., & Perner, J. (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition*, 13(1), 103–128. [https://doi.org/10.1016/0010-0277\(83\)90004-5](https://doi.org/10.1016/0010-0277(83)90004-5)
- Wolfram, S. (2002). *A new kind of science*. Wolfram Media.
- Woodward, J. (2003). *Making things happen: A theory of causal explanation*. Oxford University Press.
- Woodward, J. (2015). Interventionism and causal exclusion. *Philosophy and Phenomenological Research*, 91(2), 303–347. <https://doi.org/10.1111/phpr.12095>
- Yudkowsky, E. (2008a). GAZP vs. GLUT. *LessWrong*. <https://www.lesswrong.com/posts/k6EPphHiBH4WWYFCj/gazp-vs-glut>
- Yudkowsky, E. (2008b). The generalized anti-zombie principle. *LessWrong*. <https://www.lesswrong.com/posts/kYAuNjX2ecH2uFqZ9/the-generalized-anti-zombie-principle>
- Zheng, J., & Meister, M. (2025). The unbearable slowness of being: Why do we live at 10 bits/s? *Neuron*, 113(2), 192–204. <https://doi.org/10.1016/j.neuron.2024.11.008>
- Zuboff, A. (1990). One self: The logic of experience. *Inquiry*, 33(1), 39–68. <https://doi.org/10.1080/00201749008602210>
- Zuboff, A. (1994). What is a mind? *Midwest Studies in Philosophy*, 19(1), 183–205. <https://doi.org/10.1111/j.1475-4975.1994.tb00285.x>
- Zuboff, A. (2008). Thoughts about a solution to the mind-body problem. *Think*, 6(17–18), 159–171. <https://doi.org/10.1017/S1477175600003109>
- Zuboff, A. (2025). *Finding myself: Beyond the false boundaries of personal identity*. Philosophy Documentation Center.