**ARE QUALIA COMPUTATIONS OR SUBSTANCES?**

Mostyn Jones and Eric LaRock

Eric LaRock is an associate professor of philosophy at Oakland University and an affiliate faculty member of the Center for Consciousness Science at the University of Michigan. Email: <larockconsciousness@gmail.com>.

Mostyn Jones is a former faculty member at Washington and Jefferson College. Email: <mwj412@gmail.com>.

**ARE QUALIA COMPUTATIONS OR SUBSTANCES?**

**Abstract** Computationalism treats minds as computations. It hasn’t explained how our quite similar sensory circuits encode our quite different qualia—nor how these circuits encode the binding of the different qualia into unified perceptions. But there is growing evidence that qualia and binding come from neural electrochemical substances such as sensory detectors and the strong, continuous electromagnetic field they create. Qualia may thus be neural substances, not neural computations (though computations may still help modulate qualia). This neuroelectrical view not only avoids computationalism’s empirical issues but also its problematic metaphysics.

**1. Qualia as Computations**

**1.1 Computationalism**

The computationalist theory of mind (computationalism for short) treats minds as computational systems. Minds don’t just resemble these systems—they literally are computing systems (e.g., Rescorla, 2015).[[1]](#footnote-1) Inspired by the computer revolution, classical computationalism’s seminal ideas include that computers can think like brains and neural connections can be modeled computationally. These ideas were developed by Turing (1936), McCulloch and Pitts (1943), Newell and Simon (1956), and others. They deeply affected the cognitive and neural sciences.

Classical computationalism was based on Turing-style machines that use a table of rules to manipulate symbols on a tape in routine mechanical steps (Turing, 1936). This was combined later with the representational theory of mind, in which thinking occurs in a symbolic language of thought. These views helped create cognitive science (e.g., Fodor, 1975). Debate ensued over whether this symbolic language explains visual images (Kosslyn, 1983) and over other issues.

These computational approaches affected philosophy too. Their focus on inner representations (thoughts, percepts, etc.) superseded behaviorism’s focus on overt behavior and stimulus-response chains. Their focus on computations—which are embodied in different hardwares and thus abstracted from them—superseded mind-brain identity theory’s narrow reduction of minds to concrete (non-abstract) brain hardware. Computationalism aligns with functionalism, where pain, for example, is characterized in terms of its causes and effects, both overt and internal. Myriad computations implement each of these mental functions.

 Both computationalism and functionalism drew support from Putnam’s (1967) argument that identifying pain with brain events is implausible because this identity must hold across all species in evolutionary history. He inferred that pains are instead mental functions realizable in and abstracted from multiple materials. This helped establish today’s dominant theory of mind—nonreductive physicalism—in which pains are reduced to functions that are in turn nonreductively realized in multiple hardwares, whether organic or inorganic.[[2]](#footnote-2) Another influential argument supporting mind-as-computer views over mind-as-brain views was that no pictorial images are found in brains, so images can’t be neural events but must instead be encoded in information non-pictorially (Sperry, 1952).

A major development in computationalism was connectionism (e.g., Rumelhart et al., 1986). Unlike classical computationalism, connectionism is inspired more by neurophysiology than by computer science. Its computations are modeled on neural networks instead of Turing-style operations. These networks have hidden levels between their input and output levels, feedback loops between the hidden units, and learning algorithms that adjust the weights between the hidden units so as to train the network to yield desired outputs. Arguably, classical Turing-style computations are rule-governed symbolic manipulations that model higher-level cognition (e.g., reasoning), while connectionist computations are hidden in the weighting of connections that model lower-level cognition (e.g., learning by association).[[3]](#footnote-3)

Arguably, a further form of computationalism is computational neuroscience (e.g., Paul Churchland, 1986a; Pat Churchland, 1986). It focuses on the computations in actual brains rather than the abstract Turing-style computations of classical computationalism and the abstract, weighted-connection computations of connectionism. This sole focus on neural science to the exclusion of abstract computer science drops the multiple-realization commitment of the other two views. Ultimately, all three views account for mental operations in terms of computations, but with different views of computation. (Yet computational neuroscience may in the end not be computationalism at all.) Despite these differences, it’s often felt that these cognitive and neural approaches may ultimately be integrated to join psychological theories with neural mechanisms.

In what follows, we’ll look into four issues facing computationalist accounts of qualia (we define qualia as conscious sensory and emotional qualities such as pain or fear). (1) How are qualia seated (metaphysically) in neural activities? (2) How do the various qualia arise? (3) How do they unite to form perceptions? (4) How does their causality work?

In addressing these issues, computationalists have tried to explain how different qualia arise in terms of labeled lines (for taste, vision, etc.) and across-fiber comparisons that resolve ambiguities in labeled lines (ambiguities such as which light wavelengths are entering retinas). Similarly, different emotional qualia are explained in terms of balances and interactions between various hormonal activities. Computationalists have also tried to show how all these qualia unite to form overall experiences (of a cappuccino, a lover’s smile, etc.), how these qualia are seated in neural functions, and how their causality works. But all these accounts raise serious issues.

To flesh out these explanations and their issues, we’ll turn in the remaining parts of §1 to several prominent examples of computationalist approaches in cognitive and neural science.

**1.2 Churchland’s Neurocomputational Qualia Spaces**

Paul Churchland (1986a,b) devised reductive strategies to explain the mind. Two issues from above that he addressed were (1) how qualia are metaphysically seated in neural activities, and (2) how the various qualia arise.

 He argued that the brain represents the world via positions in computational spaces—and performs computations on these representations via “transformations of coordinations” (1986b). For example, topographically organized sensorimotor maps manage to coordinate positions in sensory space (e.g., for eye position) relative to positions in motor space (e.g., for limb position).

Computations also represent qualia. For example, colors arise from three retinal cones, which detect long, short, and medium light wavelengths. The firing-rate responses of all three cones to a particular stimulus can be represented geometrically along the three axes of a cube. The resulting color is represented by the position in the cube given by the triplet of these responses (this is an “across-fiber-pattern” account of perception). The long-wavelength response dominates for red, while medium dominates for yellow. Orange lies between them in the cube, close to the red and orange—but far from the blue.

Initially, Churchland (1981) argued that neuroscience will ultimately explain qualia fully, so qualia can be eliminated from our ontology. But, echoing Descartes, critics replied that we can always doubt that brain matter exists, but not that our own qualia exist. Churchland (1986a) shifted from eliminating qualia to reducing them. For example, while coffee appears warm to the outer senses, this warmth is actually reducible to the vibrational energy of molecules. Similarly, while inner states appear to introspection as qualia, neuroscience will show that qualia are actually neural computations. So, orange is actually a triplet of wavelength responses by retinal cones (but he didn’t explain how these computations could appear to introspection as orange).

Patricia Churchland speaks in similar reductionist terms today. For example, “there is just the brain—there is no separate thing, me, existing apart from my brain” (2013, p. 34), so “[p]robably the soul and brain are one and the same” (ibid, p. 60). She admits that minds seem to differ from brain events. But she adds that we can’t move from the fact that we don’t presently know how they can be identical to the claim that they can’t possibly be identical (p. 53ff.). She further argues that progress in physical science has led to the abandonment of dualist claims that we have nonphysical souls (pp. 44ff.). Besides, how can nonphysical minds or souls move physical bodies—what kind of energy could souls have (p. 51)?

She says that becoming conscious of anything requires activation of the brain stem, central thalamus, and cortex (pp. 233ff.). To explain consciousness of specific events, she turns to the theories of Baar and Dehaene (see below). Here, diverse brain activities are unified together into a coherent form by firing together in synchronized lockstep (p. 247), which encodes their unity (this touches on the third issue above: how do qualia unite to form perceptions). She adds that “neurons work together . . . without a conductor, without a commander in chief.” There is no “self-contained module—the will.” Instead, a “network of areas, rather widely distributed, partially overlapping, regulate self-control” (p. 176).

Patricia Churchland, unlike her husband, focuses on emotional qualia. She’s skeptical of those who simply attribute specific emotions to specific hormones (or their receptors). Instead, she links emotions to complex interactions between different hormones. For example, a mother’s defense of her brood is attributed to a mix of oxytocin, progesterone, and vasopressin. This suggests to her that the emotion of love isn’t tied simply to the oxytocin hormone, counter to some researchers’ suggestions (pp. 101, 148). But she says little about how specific emotions such as love actually arise through complex hormonal interactions (this is issue (2) above).

This approach bears some similarities to Loveheim’s (2012) hormonal approach to emotions. He assumes (like many researchers) that serotonin correlates with the emotion of self-confidence, while dopamine correlates with anticipation and motivation, and noradrenaline correlates with distress. In his view, varying levels of these three serve as the three axes of a computational space—a cube. These axes generate the cube’s eight corners, representing the emotions of anger, disgust, surprise, fear, joy, shame, excitement, and anguish—which Tomkins (1981) treated as the eight basic emotions. So, the varying levels of the three hormones are assumed to be neural correlates of all these basic emotions. These hormones help the limbic system process and disseminate emotional information—and trigger emotions in the brain.

**1.3 Dehaene’s Global Neural Workspace Theory (GNWT)**

The GNWT of Dehaene, Changeux, and others is a well-known computational theory of mind. It builds on Baars’ (1988) idea that consciousness is the central workspace of cognitive processing. Dehaene et al. (2011) cast this cognitive workspace in neural terms as neocortex’s reciprocal projections to its subareas. When ascending parallel processing inside sensory cortices reaches a threshold, this triggers a global ignition in which selected information is broadcast into the global workspace, where it’s available to memory, planning, etc. This makes the information conscious.

The global workspace consists of several interconnected hubs—the dorsolateral prefrontal cortex, anterior cingulate cortex, inferior parietal cortex, and precuneus. These hubs serve as scaffoldings for efficiently integrating information from diverse sources (this reduces the need for cumbersome amounts of interconnections between neurons). Their activation involves, for example, a P300 response (a frontal-parietal event evoked by sensory stimuli), increased gamma oscillations of neurons in the ignited areas, and synchronized thalamocortical feedbacks.

This synchronized activity unifies the hubs together in temporary coalitions as attention shifts about. Ultimately, “conscious access is global information availability” within this workspace (Dehanene et al., 2011). Indeed, “global availability of information is precisely what we subjectively experience as a conscious state” (Dehaene, 2014). This echoes the Churchlands’ reductionist language above. This account of consciousness is influential, yet it might be better construed as a theory of high-level, attentive consciousness instead of consciousness generally. For not all consciousness is attentive (Tsuchiya, N. & Koch, C., 2008).

Concerning qualia, Dehaene & Naccache (2001) argue that the problem of explaining them might be dissolved by accounting for them not just in terms of phenomenal consciousness (where qualia are experienced) but also in terms of access consciousness (where qualia are reportable and available to thinking and language). Both are facets of the same underlying phenomenon (yet they don’t say how the phenomenal facet can be accounted for in terms of the access facet). In these various ways, GNWT addresses two issues from above: (1) how qualia are seated in neural activities and (3) how they unite to form perceptions.

**1.4 Tononi’s Integrated Information Theory (IIT)**

We’ll cover IIT in more detail because of its recent prominence as a neural account of consciousness and because it addresses all four issues above: (1) how qualia are metaphysically seated in neural activities, (2) how the various qualia arise, (3) how they unite to form perceptions, and (4) how their causality works.

 IIT starts in a Cartesian spirit by acknowledging the existence of experience, then asks how brains must be organized to support it (Tononi et al., 2016, Koch, 2019; Tononi & Koch, 2015). IIT ultimately leads to a qualified computationalism.

These authors start with various axioms (Tononi et al., 2016; Koch, 2019). For example, consciousness exists intrinsically, that is, for itself, without need for anything external, such as an observer. We know it without turning to neuroscience or physics. The intrinsic nature of our experience is indubitable: it can’t be extrinsic to us instead of intrinsic, for then it wouldn’t be ours. Each experience is structured in that it contains distinctions—for example, the shapes and colors in a visual image. Each experience is informative due to all these distinctions. It’s also integrated in that its parts are a unified whole irreducible to the parts. Finally, each experience is definite, for if it were anything more or less, it’d be a different experience. In sum, these axioms say that every experience exists for itself and is structured, informative, integrated, and definite.

IIT addresses not only this mind side of the mind-body relationship but also the brain side (e.g., Tononi & Koch, 2008; Koch, 2019). This search for a bridge from the mental to the physical leads from the axioms above about experience to postulates about how physical systems are organized to have this experience. Tononi (2008) starts here by arguing that information and integration are the essential properties of our experience. For example, a photodiode simply differentiates dark from light, while human differentiations of dark from light further involve every other state they’re capable of perceiving. He concludes (p. 219) that “to generate consciousness, a physical system must be able to discriminate among a large repertoire of states (information) and it must be unified; that is, it should be doing so as a single system, one that is not decomposable into a collection of causally independent parts (integration).”

Altogether, there are five such postulates about how the causal relations of a physical system must be organized to be conscious. They correspond to the five axioms above about experience. In those axioms, every experience exists for itself, is structured, informative, integrated, and definite. Accordingly, the postulates pertain to the system’s intrinsic existence, composition, information, integration, and exclusion (Tononi & Koch, 2015; Koch, 2019).

To start with, the intrinsic existence postulate states that a conscious system must exist for itself without an observer (such as a neuroscientist observing the activity). For you to see a face, the visual stimuli have to produce a change that makes a difference to the perception’s neural substrate itself, which constitutes the system. Without these differences that make a difference to the system in itself, this system has no causal power over itself, no inner point of view, and no consciousness. If it lacks causal power, it can’t even be said to exist.

The composition postulate says that a conscious system must have structure that’s reflected in the mechanisms that compose the system. For example, a face recognition system must include nose and mouth recognition mechanisms. This structure is reflected in the connections, weights, and firing threshold of sensory neurons. The system’s integrated information (namely, Φ) is computed by looking at wiring diagrams for this information. All possible logic gates and their combinations are assessed to see how much they embody differences that make a difference to the system.

IIT represents this information geometrically (e.g., Tononi, 2008) in the form of the system’s “qualia space.” This is the space where each axis represents a possible state of the complex—a single combination of logic-gate interactions. Points along the axes are the probable efficacies of these various logic-gate combinations in the system. Arrows between the points represent information relationships between these elements. The overall set of information relationships constitutes the shape of the system’s qualia space, which in turn specifies the system’s experience. (The qualia we experience are the intrinsic nature of these information spaces.) Thus, colors are different sub-shapes of the same kind (for example, pyramids pointing in different directions)—while sounds are very different sub-shapes (such as tetrahedra). Even the simple color of blue translates into a staggeringly complex shape in qualia space, for it must be differentiated from all other colors, all other perceptions, and all other experiences generally.

The information postulate says that a mechanism contributes to experience only if it specifies differences that make a difference (that is, information) in the system itself. Again, these differentiations are highly complex and extensive. The mechanism specifies information insofar as it picks out its cause and effect in the system. The more that a mechanism’s past and future are determined by its present state, the higher the mechanism’s causal power becomes.

The integration postulate states that a conscious system must have causality that is unified and whole. This integrated cause-effect structure is irreducible to its parts, which are independent and non-interactive. This integrated information, Φ, quantifies how much the system shapes its own past and future. Φ quantifies how much the system’s causal structure changes if it’s cut up into its parts. If there’s no change, the system is fully reducible, and Φ=0.

The exclusion postulate is based on the axiom that each experience is definite. That is, if the experience were anything more or less, it would be a different experience. The exclusion postulate says that the only system that exists for itself rather than its parts is one that is maximally irreducible—one that is most irreducible and makes the most difference to itself. There will be nothing else in this system with more integrated information. This is the most irreducible part of the system. It’s called “Φmax,” though Koch (2019, p. 87) calls it the “Whole.” It’s the physical correlate of consciousness in the system.

The fundamental (or central) identity of IIT is that “Any experience is identical to the maximally irreducible cause-effect structure associated with the system in that state. The experience is identical to this structure—not to its physical substrate, its Whole” (e.g., Koch, 2019, p. 87; cf. Tononi et al., 2016, pp. 452f.). Importantly, Koch adds that while the conscious structure is a constellation of causal relations, “It is not an abstract mathematical object nor a set of numbers. It is physical. Indeed, it is the most real thing there is.” He further adds that how an experience feels is explained by different aspects of its causal structure. These distinctions bind together in the same experience via an overall binding mechanism. A well-integrated system thus has well-interconnected parts.

Given these postulates about the causal structure of conscious systems, how do they help refine neuroscience’s ongoing search for the neural correlates of consciousness (NCC)? IIT rejects two of the events that GNWT associated with consciousness above, namely, the P300 response and synchronized gamma-frequency firing by neurons. Koch et al. (2016) summarize the experimental evidence against these particular correlates.

Nor do the NCC include structures like the cerebellum, which isn’t well-integrated but is divided into independent modules. Instead, electrical activity in subsets of posterior cortical pyramidal cells is the best candidate for neural correlates of experience (at least sensory experience). Lesions to this “hot zone” produce blindness of colors, faces, motion, etc., while lesions to frontal cortex merely interfere with introspection, inhibition of emotions, and higher-order thought.

But proving any full NCC with calculations is difficult in IIT, for neural activity is too complex for any Φ-meter (or consciousness meter) to spell out integrated information. Yet IIT is linked to a practical tool for crudely doing so. This tool, the zap-and-zip method, measures the level of integration and diversity (“PCI”) of neural activity that achieves consciousness (Tononi et al., 2016; Tononi & Koch, 2015; Koch, 2019). These authors found that conscious people have a PCI of at least 0.31. This method may help show whether brain-injured patients are conscious. However, Hunt (2020) argues that PCI is a far-from-ideal measure of Φ and offers dubious support to it.[[4]](#footnote-4)

IIT can explain aberrant conscious states (Tononi & Koch, 2015). For example, cutting the corpus callosum splits the two cerebral hemispheres and splits the mind in two (these split minds have even fought over whether to button or unbutton a shirt). Here consciousness splits because the NCC is no longer integrated.

Tononi & Koch (2015) and Koch (2019) reject unqualified computationalist claims—such as GNWT’s—that minds can be computers. That is, they attack the implication that today’s computer simulations of brains can be conscious. For their feedforward architectures—where each layer is determined by and reducible to the preceding one—lack the feedbacks of human brains, which produce integrated information that exists irreducibly for itself. These simulations of brains are no more conscious inside than simulations of rainstorms are wet inside.

Yet Tononi & Koch (2015) and Koch (2019) acknowledge the possibility of future “neuromorphic” computers that simulate our brains accurately. This arguably fits IIT into computationalist claims that minds are computing systems of one sort or another. Another reason for treating IIT as computational is arguably that its causal structures are computations from top to bottom. It starts with wiring diagrams of logic gates, then computes the combinations and probabilities of these elements, and ends up with geometric qualia spaces.

**2. Evaluating Computational Qualia**

Computational accounts of qualia face serious issues with all four topics above concerning (1) how qualia are seated in brains, (2) how the various qualia arise, (3) how they unite to form perceptions, and (4) how their causality works. Each will be discussed now in this order.

**2.1 How Are Qualia Metaphysically Seated in Neural Activity?**

Computationalists treat qualia as computations. That is, qualia are reduced to computations or are their intrinsic nature. Also, they treat computations as abstract or concrete (defined as “non-abstract”). All computationalists end up in obscurity about these qualia-computation relations.

 Briefly, computationalism’s accounts of qualia-computations relations are as follows. (1) Classical forms of computationalism tend to reduce qualia to abstract, multiply realizable computations. But this is obscure about how qualia can be these abstract relations. (2) Russellian monist versions of computationalism treat qualia as the intrinsic nature of abstract, multiply realizable information structures. But this is obscure about how qualia can be the intrinsic nature of abstractions. (3) Computational neuroscience (CNS) tends to reduce qualia to concrete brain activities—not to abstract, multiply realizable computations. But this no longer treats qualia as abstract, multiply realizable computations, so it isn’t really computationalism.

 These three approaches will now be scrutinized in more detail. Arguably, specific authors may fit into more than one of these approaches.

 (1) Many classical (and connectionist) computationalists—from Putnam to Churchland—reduce qualia to computations. Here, qualia are ultimately fully explicable in terms of computations and are nothing over and above them.

 One of the earliest objections to these sorts of views is Leibniz’s argument that if we enter into the body as we do into a windmill, we’d see only figures and motions (structures and dynamics) but never anything by which to explain a perception. Mind and bodies are thus too radically different to be the same (Monadology, 1714, §17). In his New System of Nature, 1695, Leibniz adds that the mind’s subject is unified, while the body is an aggregate lacking unity.

 Reiterating Leibniz’ point that perceptions radically differ in appearance from bodies, Ewing (1962) added that we can’t find images in brains. Other critics of reductive physicalism also focus on explanatory gaps between minds and brains. For example, the qualitative, subjective nature of the mind is said to contrast with the non-qualitative, objective nature of functions and computations (e.g., Nagel, 1974; Jackson, 1986; Levine, 1993, 2021).

 Computationalists such as the Churchlands often reply that our present inability to explain how qualia and brain events are identical doesn’t imply that they aren’t identical. Nonetheless, until computationalists explain why we can’t find qualia by inspecting brains, their claim that qualia are sets of physical causal relationships will quite arguably lack intelligibility.

 There’s reason to believe that such reductive explanations will be a long time in coming. For there’s an important aspect of the explanatory gap that critics like those above don’t focus so much on.This is the gap between the concrete qualia we feel and the abstract relations of functions, computations, and information.

 This relational nature of information (for example) is obvious in standard examples of information, such as smoke giving causal information about fire, or genes giving instructive information about creating phenotypes. This relational nature is also obvious in standard theories of information, such as Shannon’s, which treats information as reductions in uncertainty, or Bateson’s (and Tononi’s), which treats information as differences that make a difference.

 The abstract nature of information structures is clear from the start. They’re formal, mathematical entities, not substantial, concrete things we can touch and feel. Similarly, coded visual images lack any color or shape. And there’s no real wind or rain in computational models of hurricanes. Indeed, information structures are abstract from top to bottom—from wiring diagrams of information flow to higher-order constructs such as integration and qualia spaces. Abstractness is actually *unavoidable* in computationalism insofar as information structures are multiply realized, for this abstracts information from hardwares.

 This abstractness is further evident from Haugeland’s (1985) argument that the blind, mechanical activities of hardwares only become meaningful information operations once we impose high-level, abstract functions on them (facial recognition, language translation, etc.). So, information operations are *necessarily* abstract, theoretical constructs in the minds of scientists.

 The overall point is that information—as well as computations and functions—are abstract relations while our sensory and emotional experiences are by contrast concrete qualities that we feel. Of course, brains are concrete too. So, inserting information alongside qualia and brains creates a third realm of abstract entities with quite obscure relations (reductive, intrinsic, causal, etc.) to the concrete realms of experience and matter-energy. (Note here that we don’t fault computationalists for using mathematical descriptions—but for giving them substantial existence and then relating them reductively, intrinsically, etc. to concrete conscious events.

 (2) Some computationalists avoid reductionism’s explanatory gaps above by treating the qualia-computation relation as nonreductive. Most notably, Russellian monists avoid reducing qualia to science’s computational and informational structures by instead treating qualia as the intrinsic, underlying ground of these extrinsic computations.[[5]](#footnote-5) Without this grounding, the world would just be abstract structure with no substance (no stuff) or qualities. One familiar example of this Russellian monism is Chalmers’ (1996, p. 305) dual-aspect theory, in which “Experience is information from the inside; physics is information from the outside.”

 But Russellian monism isn’t trouble-free. It’s arguably not just exceedingly complex but also quite obscure about how concrete qualia can be the intrinsic, underlying nature of abstract information structures. The claim that qualia ground abstract information and give it substance seems no less obscure than claims that Platonic forms are embodied in (present in) matter. Also, aspect relations (such as Chalmers’ or Spinoza’s) are deeply mysterious (Shaffer, 1968).

 One of the best-known examples of Russellian monism in neurophilosophy is IIT (for example, see Koch, 2021, p. 162). Recall its formal, mathematical accounts of qualia in terms of the shapes of causal structures. Here, the overall information relationships in a system constitute the shape of its qualia space, which specifies an experience. Colors, odors, etc. thus have different shapes. Qualia are the intrinsic nature of these extrinsic informational structures.[[6]](#footnote-6)

 IIT thus faces the issue of how qualia can be the intrinsic nature of abstract information. Also, as argued below, IIT can be criticized for not explaining much. For example, according to Tononi & Koch (2015), IIT predicts that cerebellums lack consciousness and that split brains have separate minds. They also say that IIT is supported by research on neural correlates of consciousness, consciousness detectors, and the role of inter-area brain communication in consciousness. But Cerullo (2015) replies that other theories do these same things, which challenges IIT’s explanatory power (see §4.4 below). Our ongoing argument will try to go further than Cerullo’s below. We’ll contend that IIT doesn’t have anywhere near as much explanatory power with consciousness and qualia as is usually assumed.

 (3) All the problems above in treating qualia as abstract, multiply realizable computations could be avoided by instead treating qualia as concrete brain activities. This is arguably what GNWT and other forms of computational neuroscience (CNS) are doing when they attribute (as in §1 above) qualia to neural hardware versus multiple hardwares.

 The argument that qualia are concrete brain activities would have two steps. Firstly, abstract computations are mere theoretical constructs in the minds of scientists. This aligns with Haugeland’s argument above that computations are necessarily abstract—for blind, mechanical hardware activity only produces meaningful computations once we impose high-level, abstract functions (theoretical constructs) upon it. Secondly, since it makes little sense to construe the qualia we feel as these theoretical constructs in the minds of scientists, qualia should instead be attributed to the concrete brain activities that do the real work in cognition. Importantly, this two-step argument implies that CNS isn’t a form of computationalism. For qualia aren’t attributed to computations (which are necessarily abstract) but just to brains (which are concrete).

 CNS’s problem here is that it faces an explanatory gap between qualia and concrete brain activity, just like classical computationalism faces between qualia and abstract computational activity. Whether qualia are identified (in either type or token ways) with abstract computations or concrete brains, this identity is left obscure and unexplained. Similarly, Russellian monism is obscure about how qualia can be the intrinsic nature of abstract information.

 Our approach to qualia below will resemble the non-computationalist CNS above. We’ll treat qualia as concrete brain substances while treating computations as mere theoretical constructs in the minds of scientists. (Here, concrete will mean non-abstract, and substance will mean the concrete stuff comprising brain activity.) We’ll try to avoid CNS’s explanatory gap by identifying qualia with these substances in nonreductive ways.

 Below, the term computations will refer to abstract functions that exist only as theoretical constructs in scientists’ minds. The term “computations” (in quotes) will refer to the concrete brain activities that do the real work in cognition. The latter are in quotes, for strictly speaking they aren’t true computations (true ones are necessarily abstract, as Haugeland noted).

**2.2 How Do the Various Qualia Arise?**

Most of the computational views above address the origins of our various qualia. Yet each raises its own problems, leaving these origins quite problematic in computationalism.

To start with, neuroscientists now explain how our different sensory qualia arise in terms of specialized labeled lines with their own detector fibers and processing areas for taste, vision, and other sensory modes (e.g., Purves et al., 2001; Parker, 2019). In this way, photoreceptors produce color qualia regardless of whether they’re stimulated by light, pressure, or other stimuli. This process is supplemented by detailed comparisons of the fibers within each labeled line (ibid.). For example, the three color fibers overlap in their response to long, medium, and short wavelengths of light. So across-fiber comparisons of their firing help disambiguate which wavelengths are actually present (e.g., Conway, 2009; Solomon & Lennie, 2007).

 Paul Churchland’s account of qualia above is limited because it focuses just on the across-fiber comparisons. These comparisons alone can’t distinguish colors from sounds (for example) because across-fiber comparisons operate within—not between—sensory modes. Moreover, in this across-fiber method, all fibers are treated as inherently the same—they differ only in their firing patterns. Yet there’s no evidence available that these patterns actually differ in systematic ways between the modes. So, this method again fails to distinguish colors from sounds. This is why across-fiber comparisons are seen as mere supplements to labeled lines.

Now neuroscience’s basic problem here is that this combined method is so similar among the visual, tactile, and other sensory modalities that it’s unclear how these methods can actually differ enough to account for all the stark differences between qualia. We’ll argue in §4 that sensory qualia come not from the computations of sensory-detector lines but from the electrochemical substance of detector cells, such as the opsin proteins in retinas. This resembles Hering’s (1913) idea that different sensory qualities come from different modes of nerve energy.

The point isn’t that “computations” play no role in brains. For “computations” (with the quote marks indicating just concrete neural interactions, not abstract functions) do undoubtedly operate pervasively behind the scenes to refine images—using mechanisms for constancy, depth, disambiguation, saccadic suppression, object recognition, et cetera. The point is instead that these “computations” don’t explain how the different qualia arise. “Computations” establish activity levels of certain fibers over others, but there’s growing evidence for the testable claim that EM activities in the fibers are what ultimately establish the qualia of these fibers.

Turning to IIT, its accounts of qualia spaces are far too complex to specify, let alone to test—they remain purely speculative. For example, a simple patch of blue color translates into a highly complicated shape in computational space. This patch must be differentiated in terms of its contrast to myriad other colors, shapes, locations, durations, and movements. Further, it must be differentiated in terms of, for example, its visual versus auditory nature, its topographical versus categorial nature, and its sensory versus nonsensory nature. The combinations are staggering (cf. Aaronson, 2014).

A useful, viable theory of how different qualia arise needs to spell out the neural correlates of qualia in testable ways. But IIT’s resistance to testing—and its reliance on axioms, thought experiments, and abstract mathematical accounts— ultimately make it seem less like an empirical hypothesis than a rationalist speculation.

Turning from sensory to emotional qualia, Patricia Churchland says (as already noted) that the physiological functions of hormones are too numerous and complex to equate, for example, oxytocin’s function with simply being the love molecule. She instead ascribes emotions to complex hormonal interactions. But she doesn’t specify any of the interactions. Nor does she say why one of the many functions of oxytocin can’t be to serve as the seat of conscious love.

Despite her skepticism, there’s considerable evidence for correlating oxytocin with maternal and romantic love. For example, without a key oxytocin-activated pathway, maternal behavior fails to develop in mice (Busnelli & Chini, 2017). Also, oxytocin antagonists block mating-induced pair-bonding (Ross & Young, 2009), and intracerebroventricular infusions of oxytocin induce partner preference in un-mated females (Insel & Hulian, 1995).

Churchland claims that oxytocin could serve as an anti-anxiety hormone instead of the love hormone. But these two roles don’t really preclude each other, for feelings of love can go long ways in soothing anxieties between individuals. In the end, she gives no evidence for where love comes from and scant evidence that it doesn’t come from oxytocin.

Lovheim offers a highly debatable computationalist account of emotions too. His correlations of serotonin with self-confidence, noradrenaline with distress, and dopamine with anticipation are widely accepted. But there’s far less evidence for his systematic correlations of other emotions with *varying mixtures* of these three hormones.[[7]](#footnote-7) (It’s quite likely that these hormonal mixtures instead just affect the other emotions indirectly by simply modulating the levels of these emotions’ own corresponding hormones—after all, it’s widely known that hormones do modulate each other’s levels.) So, it remains quite likely that emotional qualia instead *correlate closely, even one-to-one, with their own hormones* like serotonin does with self-confidence. In the end, there’s scant evidence that hormone interactions determine specific qualia, yet it’s widely accepted that these interactions do modulate emotional intensity levels.

So, computationalists don’t adequately account for how our various sensory and emotional qualia arise. We’ll argue below that these qualia don’t arise from computations but from the electrochemical substance of sensory and hormonal detector cells.

**2.3 How Do Qualia Unite to Form Perceptions?**

IIT’s integration postulate says that a conscious system must have unified causality. But IIT doesn’t explain how brains actually produce unified causality. At issue here is neuroscience’s unresolved binding problem, which concerns how dispersed neural circuitries create a unified experience (for example, the unified sensory and emotional experience of seeing an old friend) and how they create the unified mental subject who owns them. This harks back to Leibniz’ point that the subject is unified, while the body is an aggregate lacking unity. So, how does the brain encode the obvious unity in our experience? This isn’t explained by the standard candidates of synapses, feedbacks, neural synchrony, or attention.

To start with, Zeki (1993, p. 296) reported that “there is no single cortical area to which all other cortical areas report exclusively, either in the visual or in any other system” (cf. Zeki, 2003, 2015). More recent studies still find no evidence that a single central brain circuit binds cortical activity into unified, coordinated experience (Jones, 2019). This isn’t surprising, given the combinatorial capacity limits of brain circuits (Crick & Koch 2003; LaRock, 2007; Zeki 2003).

 This binding (or unity) problem can be illustrated in more detail with visual images. To start with, some ventral-cortical detectors integrate many lower detectors to recognize particular objects, such as faces, as unified wholes. Yet there are no top-level detectors to recognize all possible visual scenes. Indeed, can never have a top-level detector for each possible visual scene. So, while standard neuroscience has explained our vision of some shapes and objects, it has not yet explained our perception of the overall unified shapes and layouts in visual images.

 Also, instead of using a single, central processing area to unite colors and shapes, vision seems to use separate, parallel circuits. So how do they bind together to form complete images? Zeki (2003) argues that ascending color and shape circuits have few (if any) synapses to link their cells. Arguably, he overlooks how higher feedbacks into lower cortical maps might indirectly bind color and shape (Kawato, 1997). But to encode detailed images, these feedbacks would have to systematically connect all shape and color features point by point across the whole visual field—thus achieving what even highly detailed lower cortex fails to do. All these various problems above have foiled connectionist efforts to explain cognition via neural connections.

An alternative view is that the synchronous firing of neurons binds them into a unified, conscious form (e.g., Gray et al., 1989). Patricia Churchland and GNWT expressed this view above. But Koch points out that binding can occur without neural synchrony (this is even shown by Gray and Singer’s (1989) own data, which shows that color and shape circuits don’t fire in neural synchrony during perception). Also, this synchrony occurs without binding or conscious unity during seizures, anesthetized states, and NREM sleep (Koch et al., 2016; LaRock, 2019).

Finally, Crick and Koch (1990) once held that focal attention binds perceptual features into a unified, conscious object (as when we suddenly spot a friend in a crowd). But binding can occur without attention. Treisman (2003) notes that subjects reported mistaken combinations of color and shape in objects if their attention was diverted—and this involved binding, despite the mistakes (LaRock, 2007).

Given that standard binding theories are so problematic, it’s understandable that IIT fails to explain how brains actually produce the unified causal structures that IIT’s integration postulate stipulates. Nonetheless, this failure casts further doubt on how useful IIT is as a theory of mind. Below, we’ll argue that this unity comes not from neural computations but from the continuous, unified substance of the neural electromagnetic field.

**2.4 How Does Qualia Causality Work?**

The various computationalist accounts of qualia’s causal properties are problematic. One account is the Churchland’s reduction of qualia and their causality to brains and their causality. This renders qualia into causally impotent epiphenomena of brain events. This view is hard to refute, yet there’s no actual evidence that it applies to higher cognition and thought, and it’s highly implausible on evolutionary grounds (Jones, 2017).

 Some computationalists do attribute causal powers to qualia. For example, Tononi (2015) rejects causal reductionist claims that only first-order elements (such as atoms) exist, and all causality comes from their interactions. According to IIT’s intrinsic existence postulate, a system’s existence depends on it having a maximum causal power. A system with this power exists at a second order above any fundamental, first-order elements, and its activities introduce second-order effects over first-order effects. So, mental causes are irreducible to neural causes.

A similar but more developed computationalist approach to mental causation comes from Murphy and Brown (2007). They too reject the view that systems control their components in the mechanistic, bottom-up ways of causal reductionism. In their systems approach, self-controlling systems aren’t concrete things, they’re instead hierarchical information-processing structures (p. 87), and this causality is usually not mechanistic and deterministic but a matrix of probabilities running on information (like they attribute to the flight of a fly on pp. 77f., 97f.). The basic contrast here is between bottom-up causation, where the parts mechanistically determine the whole—and top-down causation, where (in their view) the whole exists over and above the parts and introduces its own autonomous causes that affect the parts.

The problem here is that these irreducible causal structures add a new level of causality that’s just a theoretical construct, not a real thing (cf. Haugeland above). For example, Bunge (1979, pp. 13-14) argues that wholes can’t affect parts—a level of organization “is not a thing but a set and therefore a concept . . . All talk of interlevel action is elliptical or metaphorical.”

It’s hard to see how abstract theoretical constructs can make matter do anything. These abstractions can’t push ion currents about in membranes or pull neurotransmitters across synaptic clefts. This is much the same concern that Plato raised when he said ideal forms affect the world of matter. For the same reason that it was hard to identify abstract information with qualia above, it’s also hard here to posit causality between abstract information and matter.

This criticism could be avoided by an ontology of pure information in which the universe (including matter and mind) is nothing but information. But Chalmers (1996), Rosenberg (2004), and others reply that this is a world without substance—a sheer abstraction, not a real world.

So computationalist accounts of qualia causation face problems. These problems—and those above concerning how the various qualia arise, unite, and are seated in brains—make it hard to embrace assumptions that scientific progress shows that qualia can ultimately be explained in scientific, computational terms. We’ll now offer a qualia theory that opposes computational accounts yet draws on considerable recent experimental evidence.

**3. Qualia as Substances**

**3.1 EM-Field Theories of Mind**

The noncomputational theories below try to avoid the computationalist issues above about how qualia are seated in brains, how their varieties arise, how they unite to form images, and how their causality works. These theories do so by seating subjects and their qualia not in neural computations but in neural substances—specifically, in neural electromagnetic (EM) fields and the electrical charges and currents generating them. Here, substances are treated as the concrete**,** fundamental kinds of stuff in space-time that comprise the universe—most notably, EM.[[8]](#footnote-8) This reflects the word’s Latin roots in *substantia* (foundation)—and perhaps in some ways, Aristotle’s *Categories,* where substances are enduring, fundamental things that exist independently.

 EM-field theories of mind have long seated subjects in neural EM fields. They were inspired by renowned thinkers such as Kohler, Libet, Eccles, and Popper, but their most developed forms came from Pockett, Fingelkurts, and McFadden (see Jones, 2013, for references and accounts). They’re proliferating because they draw on considerable experimental evidence, withstand past criticisms, and help to avoid neuroscience’s serious problems.

 We’ve laid out evidence above against computational accounts of minds. And we’ll layout out evidence below for field accounts of minds. But for now, to get an initial sense of the plausibility of field accounts, let’s quickly consider some similarities between EM fields and visual images. Images pictorially resemble the standing EM fields in retinas–and in brain areas mapped onto retinas. Also, both images and fields seem rather incorporeal. Further, both arguably arise from discrete neurons and reach across space in a continuous, unified form.

We’ll start with our own two nonreductive EM-field theories of mind. One may appeal to monists, the other to dualists. We’ll end by contrasting these theories to computationalist ones. The fast-growing evidence for our field theories will be listed, and tests of them will be framed.

**3.2 Our Monist Field Theory**

Our monist field theory is best characterized relative to traditional (Cartesian-leaning) dualism, where subjects are nonspatial substances causally linked to their physical, bodily substances. This dualism faces the issue of how radically different subjects and brains can interact. Lindahl & Arhem (1994) have mounted a sophisticated defense of classical, Cartesian-leaning field theories—from Kohler, Libet, Eccles, and Popper—that seems less vulnerable to this traditional dualist issue. They argue that the subjective mind is nonphysical yet still interacts with the brain’s action-potential patterns via the mediation of the brain’s EM field.

In contrast, monist field theory (MFT) offers a view in which only one substance exists. This non-reductive monism draws on realist claims that we experience our thoughts and feelings directly—while perceiving matter quite indirectly via eyes, instruments, reflected light, etc. So, we can’t know matter’s underlying, hidden reality beyond its sensory appearances (in Lockean terms). Authors from Russell onward have added that we can’t know what brain matter is really like beyond perceptions of it. So, for all we know, this underlying reality of the brain could be the mind. Strawson (2016) adds that all matter-energy may be conscious like this (panpsychism). Physicists cannot object to this view, for they describe all particles and fields solely by their observable effects—while MFT refers to what particles and fields are in themselves, apart from their observable effects.

Our MFT adds to this familiar type of panpsychism that while even atoms have a minimal consciousness, the developed subject (or mind) and its qualia are the underlying reality of the brain’s electrochemical activity. The EM field in this activity binds the activity into a unified conscious whole (all energy fields are conscious, but only EM is powerful along brain circuits). This conscious whole disintegrates back to molecular levels as neuroelectrical activity wanes. Consciousness is then atomized, as it were (see §4.2 below).

MFT’s virtue is its simplicity and clarity relative to other mind-body theories (Jones, 2016a). It’s an ontological monism, for just one substance exists—namely, consciousness which occupies space and exerts forces. This monism doesn’t reduce conscious activity to neuroscience’s observable activities (as in the Churchlands’ intertheoretic reductions). Instead, the conscious subject and its qualia are the underlying reality of neuroelectrical activity *beyond* neuroscience’s descriptions of it. This avoids Levine’s explanatory gap (unlike the computational accounts above). It also offers a reply to Leibniz’s argument—the reason we can’t find qualia as we inspect the brain is that they reside in the brain beyond what we perceive of it. Our monism also avoids the problematic abstract computations that nonreductive theories (including Russellian monist, computationalist, and functionalist ones) rely on.

Our MFT thus stresses that reality is a single substance.[[9]](#footnote-9) Despite this ontological monism, MFT resembles dualism in a key way. For MFT isn’t just ontologically non-reductionist (as explained above) but also causally non-reductionist. This is, in effect, a causal dualism. Two examples help make this non-reductive causality clear.

First, while neuroscientists can detect an EM field shifting across a brain, they can’t show how the field weighs moral choices or even chooses which foods taste best. For example, if we compare wines, our ranking of them depends on brain activities used in perception, memory, et cetera. Yet the crux of the decision comes from intuitive, conscious qualia comparisons. In MFT, it’s these conscious comparisons in the unified field that makes the choices, not the mechanics of neural circuits. Epiphenomenalists can’t argue that these conscious choices (and consciousness generally) are impotent here, for all causality is conscious in MFT. Even putting MFT aside, it hardly seems plausible to argue now that these comparisons wholly depend on the mechanics of neuroscience—or the electrodynamics of physics. For these disciplines haven’t shown how qualia are consciously chosen. Until physics clearly and precisely explains this choice of wines, causal reductionism seems little more plausible than ontological reductionism. For example, if we isolated the precise EM field patterns for wine qualia, how exactly how would we go about explaining *the actual choices of which wines taste best* in electrodynamical terms?

A second example that shows how MFT (and DFT below) have non-reductive causality comes from our ability to rewire our brains after strokes. In constraint-induced movement therapy (CIMT), the healthy limb is constrained, so the patient must use the injured limb. This injured limb improves in patients who exert considerable mental effort in the therapy, indicating that *self-directed neuroplasticity occurs* (LaRock et al., 2020). These CIMT and wine-tasting examples show that mental causality in MFT and DFT has emergent dynamics that go beyond physics. Such points thus help show how MFT and DFT’s emergent causality can counter the ontological and causal reductionism in Churchland’s view that minds are just brains.

We realize that many people instead think that mental dynamics supervene on the dynamics of neurons, EM, et cetera. But while there’s evidence for this supervenience in simple cognition, there’s no evidence for it in deliberative thought (Pockett, 2006). There’s plenty of introspective evidence that we make choices by comparing competing feelings about situations, yet there’s no scientific evidence that these choices are instead made by physical activities governed by physical laws. Again, introspection shows that we choose which wines taste best by comparing their qualia, but there’s no evidence from biophysics that weighing these qualia is nothing but shifting neural EM patterns governed wholly by electrodynamics principles. (Jones, 2017 and Jones & LaRock, 2021 for longer arguments.) Our points here are not that supervenience cannot apply to mental causation, but just that, firstly, there’s no conclusive scientific evidence for the assumption that supervenience does apply to all mental causation, and secondly, it’s hard to see how supervenience can apply in our theories.

**3.3 Our Dualist Field Theory**

Our alternative to MFT is a dualist field theory (DFT). Both have their virtues. The latter is best depicted relative to Lowe’s (2010) Non-Cartesian Substance[[10]](#footnote-10) Dualism (NCSD). In NCSD, subjects and their causality are emergent from brains and irreducible to neural terms. Yet subjects have spatial properties that allow them to interact with brains without Descartes’ causal issues. Subjects are also simple, unified substances, unlike brains’ myriad, dispersed neurons.

Our DFT is a modified form of Lowe’s NCSD (Jones & LaRock, 2021). It addresses two issues Lowe wasn’t clear about. First, Lowe treated subjects as *unified, simple substances*. But he wasn’t clear about how subjects come to be this way. Second, Lowe insisted that subjects don’t *exert forces*. But he wasn’t clear about how the conscious purposes of subjects can affect bodily behavior without having any forces or energy transfers.

DFT’s modification of NCSD addresses these two issues by turning to EM field theories of mind. First, DFT agrees with MFT that conscious subjects are the underlying reality of neural EM fields. An EM field is thus a *simple, unified substance*, unlike the separate neurons that generate it. It’s unified because it reaches across various neuronal spaces and neuronal processing times as a continuous particle-wave substance (quanta in strong fields form a probability cloud of continuously high energy). It’s simple in that it’s a continuous, unified whole relative to the separate, discrete neurons it comes from. Second, in both DFT and MFT, the subject *exerts EM forces*, interacts with brains, and is causally and ontologically irreducible to the observable events of physics. The subject exerts forces because it’s part of the underlying nature of EM fields and thus occupies their space and exerts their forces. (This counters Patricia Churchland’s criticisms above of dualism and its causality.) But DFT diverges from MFT’s monism—instead, two substances exist: the conscious substance of EM fields and the nonconscious substance of all particles and all fields except EM fields.[[11]](#footnote-11)

DFT shares MFT’s causal dualism, where conscious, intentional causation interacts with blind physical causation. But DFT trades MFT’s substance monism for a substance dualism in which everything is nonconscious except for the conscious substance of EM fields, which exist in space as a form of energy, and constitutes (in brains) the mind and its subject..[[12]](#footnote-12)

 To summarize, both MFT and DFT are causal dualisms where the causal dynamics of minds and brains differ radically yet interact. But these theories differ ontologically. MFT is ontological monism, for consciousness is the underlying, hidden reality of all matter-energy beyond how it appears to the senses. DFT is ontological dualism, for consciousness is the underlying reality of just EM fields, while all else is nonconscious.

 This runs counter to the often-held principle that the physical domain is causally closed in that physical events are fully explicable by laws of physics. But until neuroscience is advanced enough to explain qualia, this principle is hardly compelling. Indeed, if the arguments above are right, physical science’s computations, structures, and dynamics will never explain qualia.

**4. Evaluating Substantial Qualia**

We’ll now show how MFT and DFT try to deal with the four key issues concerning qualia, which computationalism fails to explain.

**4.1 How Are Qualia Metaphysically Seated in Neural Activity?**

As already indicated, our MFT and DFT offer ways to avoid issues with physicalist reductions, Russellian monist structuralism, computationalist abstractions, and dualist causality. We don’t reduce qualia to the observable EM energy of physics. Qualia are instead the underlying nature of EM energy beyond physics’ description of them. We also avoid Russellian monism’s issue with how qualia can be the “intrinsic nature” of abstract information structures. Again, qualia are just the hidden substance of EM behind its observable behavior—there are no abstract structures. Finally, while we agree that “computations” play big roles in cognition, we treat them not as the abstract causes of some computationalists—nor as the nonspatial causes of traditional dualists—but just as the underlying nature of concrete electrochemical processes. In these ways, MFT and DFT try to offer clear, simple mind-brain body theories that avoid deep, perennial metaphysical obscurities. Yet their neural details below will inevitably involve some complexities.

**4.2 How Do the Various Qualia Arise?**

Seating our qualia in the brain’s electrical activity allows our field theories to offer a testable explanation for how different qualia arise. This enables these theories to avoid computational neuroscience’s problems here—for example, that processing circuits don’t differ enough to explain the stark differences between qualia in the visual, taste, and other modes.[[13]](#footnote-13)

In attributing qualia to neuroelectrical activity (its charges and the fields they generate), our theories draw on fast-growing evidence that different sensory qualia correlate with intense, localized electrical activity in different sensory-detector cells. The electrical activity occurs in the cells’ membranes, namely in their unique ion-channel proteins and G-protein-coupled receptors (GPCRs). These electrically charged proteins are vital neuroelectrical activities—they generate nerve impulses and fields. In our theories, qualia aren’t encoded by these proteins. Qualia are instead the underlying substance of the proteins’ charges and fields—the stuff they consist of beyond their sensory appearances (this is why qualia aren’t observed in brains). Fuller arguments for this overall view appear in Jones (2019)[[14]](#footnote-14)

Examples of these qualia-protein correlations are that temperatures correlate with unique ion channels, such as TRPM8 for cold, TRPV3 for warm, and TRPV1 for hot. Also, sweet tastes correlate with the unique GPCR complex T1R2/T1R3, savory correlates with the T1R1/T1R3 complex, and bitter correlates with the GPCR family TAS2R (https://www.genecards.org/). Additionally, the primary colors correlate with OPN1MW, OPN1LW, and OPN1SW, which are GPCRs of the opsin class (ibid.).[[15]](#footnote-15) Reflecting a general principle with these proteins, the three opsins form the three labeled lines for colors, while across-line comparisons modulate which of the three is most active (e.g., short-wavelength light strongly activates the blue line and inhibits the other two lines). These various proteins tend to reside from peripheral to cortical levels.

There are far too many correlations across the various sense modalities to cover here (see Jones, 2019, for a much fuller list with citations).[[16]](#footnote-16) There’s also evidence that such correlations hold between emotional qualia and limbic hormonal receptors. For example, oxytocin and vasopressin proteins correlate with feelings of love (as argued above), while estrogen and testosterone correlate with lust (Fisher, 1996), endorphin correlates with euphoria (Sprouse-Blum et al., 2010), and adrenaline correlates with vigilance (e.g., Bayerl & Bosck, 2018).

We go beyond this evidence that (for example) oxytocin correlates with love—to actually attribute love to oxytocin receptors. This does not oversimplify the complexities of love, for this attribution is only meant to account for the innate emotional feeling (sheen) of love experiences, not their crucial conceptual components, which are highly complex and learned. Also, levels of oxytocin are affected by interactions with other hormones, so feelings of love are modulated in complex ways by other hormones. Finally, note that scientists were shocked by the discovery that genetically altered prairie voles who lack oxytocin still pair bond normally (Berendzen et al., 2023). But while this study shows that oxytocin is not necessary for pair bonding, it does not show that oxytocin is not necessary for feeling love. Indeed, the authors argue that these altered voles may well have compensated for their lack of oxytocin by activating vasopressin pathways which preserve this vital pair-bonding behavior to safeguard the species’ survival.

It may seem that neuroplasticity threatens this account of qualia. For example, if visual cortex is recruited for somatosensory processing by blind subjects, and these cortical detectors are stimulated, then subjects report somatic qualia (Ptito et al., 2008). This might threaten our claim that visual-detector proteins correlate with visual qualia. In reply, our view isn’t threatened if neurogenesis and plasticity yield not only new detector synapses, but also new detector GPCRs and channels. Many somatosensory GPCRs and channels already exist in occipital and parietal lobes (Su et al., 2004), so neurogenesis of more of them would hardly be surprising.

If all our different qualia ultimately do correlate with different electrically active proteins in detector cells, this would support our neuroelectrical approach over existing theories of how different qualia arise. For existing theories attribute qualia and their binding to computations by overall circuits, not to neuroelectrical substances in cell membranes. So, whether these qualia-protein correlations continue growing constitutes a test of our theories.

This test of whether further proteins actually do correlate with further qualia involves standard experimental procedures already in effect (see Jones, 2019 and the examples in §1.3 of this current paper, as well as the genecards.org citations above). These include establishing that certain qualia and certain neural proteins appear together in various species in evolutionarily conserved ways; that loss of the qualia (as reflected in behavior) accompanies deactivation of the proteins; and that additions of the proteins to brains accompany the qualia’s appearance (again, as reflected in behavior).

So, unlike existing IIT or EM-field theories of qualia, our theory of qualia is relatively simple and relatively testable. We identify qualia with well-known, straightforward protein activities, not with highly complex and largely unknown computational spaces or field patterns.

**4.3 How Do Qualia Unite to Form Perceptions?**

Computational neuroscience has trouble explaining how separate qualia circuits bind together to form the colors, textures, taste, etc. of a peach, for example. This binding isn’t explicable by synapses, neural synchrony, or attention. But in our theories, neuroelectrical substances (fields and charges) not only contain qualia, but also unite them to form overall percepts and emotions.

 Our two field theories explain binding via EM fields. To start with, as brain circuits activate, they generate a continuous EM field between neurons that pools their separate, “atomized” consciousness. Fully conscious experiences, such as visual images, arise when a strong field fully unifies consciousness along brain circuits. The field is strongest where ion currents rush through ion channels (and over nearby electrically active GPCRs), then run along and even between neuronal circuits (Hales, 2014). The field is especially strong among well-aligned cells (such as in cortical modules) that fire together in a coherent, mutually reinforcing manner. Yet this field degrades exponentially with distance, which can explain why consciousness is not united between brains. Even within each brain, the field is at times too weak to fully unify consciousness, leaving much brain activity merely subliminal.

 These fields help our two field theories account for binding. For example, while the brain seems to lack a single central circuitry to bind colors and shapes together, its separate circuits still generate a single, continuous, unified EM field. So, seating qualia in EM could help explain how separate colors and shapes become the unified colors and shapes in images. Even if circuits don’t connect synaptically, they can still unite if the local fields within their diffuse ion currents make contact, as color and shape circuits do in cortical maps (such as primary visual cortex).[[17]](#footnote-17)

Note that this approach to binding by EM fields contrasts with the flawed idea of binding by synchronized firing in circuits in §2.2.[[18]](#footnote-18) The former instead involves binding by coherent, synchronized fields (as in the General Resonance Theory of Hunt & Schooler, 2019). Here percepts bind into unified wholes via a strong, coherent EM field. This field is strong because the peaks and troughs of its oscillations are in phase and thus reinforce each other (versus canceling each other out). This approach aligns with arguments that critical phase transitions in neural electrical activity correlate with consciousness (Toker et al., 2021; Hunt & Schooler, 2017).[[19]](#footnote-19)

These coherent fields arise in a cycle of EM activity. Frohlich and McCormick (2010) found that fields from cortical neurons can enhance or suppress neuronal firing. This loop thereby “modulates and guides neuronal circuit activity.” This can spread synchronized firing and coherent fields over brains. In our theories, these strong, coherent fields can unify consciousness across brains. This loop can help unify mental operations. Bastos et al. (2018) found that ascending gamma-frequency oscillations are linked to sensory activity while descending beta oscillations are linked to attention’s inhibition of the gamma activity. Beta rhythms thus seem to help sculpt the focus of attention and content of working memory. In our theories, these coherent neuronal oscillations produce coherent EM fields with conscious unified perception, attention, and working memory.

As EM fields bind qualia from proteins to form overall images, the pictorial form of these images comes from sensory maps. These maps include retinas and the cortical maps they electrically connect into point-by-point. Mental images come from the undistorted visual arrays of grid and place cells in the entorhinal cortex and hippocampus, which help us navigate and imagine situations. Images aren’t observable in these areas—they’re the hidden, underlying reality of electrical activity here—beyond what’s observable of them via EEGs. In these ways, images aren’t abstract computational relations—they’re EM activity spread across neural maps.

We claim that strong, localized EM fields unify qualia from myriad proteins to form images. This can be tested by altering a proposal in Libet (1993). He suggested that the mind is a conscious mental field. This field is a unified whole that interacts with discrete neurons. It may emerge from neuroelectrical activity but isn’t reducible to it, and it isn’t detectable by any instruments. This conscious field doesn’t arise across “substantial gaps of space” or “nonneural barriers” (1993, pp. 393–4, 396), such as split neural hemispheres in brains. We have (in effect) modified this theory by treating this conscious field as the underlying, hidden nature of the brain’s EM field. This field only interacts with the brain at short distances where it’s strong.

Libet devised a test of his theory in which a slab of human sensory cortex that’s being therapeutically removed would first be isolated from the surrounding cortex (by, for example, precisely slicing connections into it with a laser). Libet predicted that if this isolated slab is artificially stimulated, it will continue contributing to reportable experiences along with the surrounding cortex. For experience doesn’t reside in neural connections but in the conscious field arising from the cortex as a unified whole (1993, pp. 396, 400).

We suggest that the test subject should be a mouse trained to respond with leg motions when three blue spots appear in a row to one eye (the other eye is covered). Its head and eyes would be held in place while a cortical slab corresponding to one of the blue spots is sliced out by laser. Our prediction is that the mouse will continue responding to the three spots, but only when the slice is under approximately 400 µm thick. (This distance is derived from Chiang et al., 2019, who discovered that EM fields—not synapses—propagate signals across sliced rat hippocampus tissue up to 400 µm.)

We’ll end this section on how EM fields unite qualia to form perceptions by now turning to some indications of how EM fields can bind further experiences—beyond just perceptions—into a unified form (for details, see Jones & LaRock, 2019; Jones, 2016b). There’s evidence (see below) that resonating EM fields (with gamma, theta, and other frequencies) help unify and guide attentive activities, working memory, imagination, and centralized executive activities.

In the latter case of the mind’s executive (or subject), there’s no evidence of a single, unified cortical area where all these conscious processes occur together (Zeki, 2003). So the best candidate for the mind’s single, unified subject is the brain’s single, unified EM field. This counters Churchland’s skepticism above about the existence of the subject. Note that the self doesn’t passively reside in this field. Its active, shifting attention binds diverse perceptions, feelings, and thoughts together. These experiences are unified by and inherent in the subject (Jones, M. & Larock, E., 2019). This field account of mental unity addresses Leibniz’ concern about how subjects come to be unified while their bodies are mere aggregates without unity.[[20]](#footnote-20)

Evidence that unified cognition comes from EM takes four forms. First, as already argued, no other mechanisms explain the mind’s unity. Second, Koch et al. (2016) argue that locally activated EEGs actually track conscious perceptions across brains better than other events, such as neural synchrony or P300 events. This EEG evidence correlates unified perceptions with local neuroelectrical fields, which aligns with our field theories. Third, EM fields alone—not any particles or synapses—propagate signals across slices in hippocampal tissue (Chiang et al., 2019). This indicates that it is the fields that unify this activity. Fourth, as noted below, there’s growing evidence that oscillating fields help our conscious attention to control cognition. This indicates that subjects may exert forces in the form of EM fields.

Again, the argument above isn’t that “computations” and synaptic connections play no roles in cognition. The argument is instead that “computations” help produce the mind’s intelligence, while neuroelectrical activity produces the mind’s qualia and unity.

**4.4 How Does Qualia Causality Work?**

Our MFT and DFT avoid classical dualist claims that bodies interact with nonspatial minds. For while minds and bodies do interact, both exist in space. We also avoid computationalist claims that abstract mental computations affect concrete neurons. For while minds do affect neurons, minds aren’t abstract. Finally, we avoid the causal reductionism and epiphenomenalism of some computationalists, such as the Churchlands. In our field theories, the conscious subject and its qualia influence neurons in a recurrent (top-down) manner via EM fields. These conscious fields aren’t causally impotent by-products of neurons, for fields may act back on the neurons generating them, as classical field theorists have long claimed (e.g., Lindahl & Arhem, 1994). This claim is based on all our arguments above that EM fields are conscious—and on the arguments directly below that (1) oscillating EM fields can make brain circuits oscillate, and (2) these oscillating circuits may in turn help guide cognition.

 (1) The prevailing view among neuroscientists, until recently, was that EM fields are so weak in brains that they’re virtually negligible—the real work in brains is done instead by neurons and synapses. But this view has changed. A series of experiments—including Frohlich and McCormick (2010), Anastassiou et al. (2011), and Anastassiou and Koch (2015)—showed that even weak exogenous (externally caused) fields can be applied to entrain spikes within slices of neural tissue. Subsequent studies showed that endogenous (internally caused) fields have this same effect. For example, Zhang et al. (2019) presented evidence that ephaptic modulations (i.e., fields affecting neuronal firings) occur in sensory circuits when alternative connections (synapses and gap-junctions) were lacking (also see Gourdie, 2019; Martinez-Banaclocha, 2020).

 Even more dramatic evidence came from 2019 studies by Durand’s laboratory at Case Western Reserve University. They showed that when a hippocampus memory circuit is sliced through to block transmissions by synapses and gap-junctions, then nerve impulses on one side of the slice still generate an EM field that actually propagates the signal across the slice, at least until the slice reaches 400 μm (e.g., Chiang, et al. 2019). Fields thus propagate signals up to 400 μm even if gap junctions and synapses are blocked. This work offers evidence that fields alone—not just particles or neural structures—effectively propagate neural activity along circuits.

 This is a sample of the evidence for our first argument above that oscillating EM fields can cause oscillations in brain circuits. Now we’ll turn our second argument that these oscillating circuits may in turn help guide cognition—thus countering claims that fields are epiphenomenal.

 (2) The frequencies (gamma, beta, etc.) of naturally oscillating neural fields come from the firing frequency of the neurons generating the fields. The topic of how these various frequencies affect cognition is a huge, fast-growing area of research. But we can at least very briefly list some directions it’s going in and some of the researchers involved. One thing these studies indicate is that oscillating fields affect circuits by sculpting the focus of attention and content of working memory. For example, as already noted, gamma-frequency oscillations correlate with bottom-up sensory activity, while beta oscillations correlate with attention’s top-down inhibition of the gamma activity (Vernet et al., 2019; Bastos et al., 2018; Narikiyo et al., 2020). Such oscillating circuits also help guide long-term memory (Lisman et al., 2013; Heusser et al., 2016; Ezzyat et al., 2018; Meyer et al., 2017), working memory (Bahramisharif et al., 2018; Riddle et al., 2020), imagination (Schafer and Schiller, 2018; Kay et al., 2020).

The last of these studies is a particularly striking example of how oscillating fields may help promote cognition. Kay et al. showed that different hippocampal place cells (which are used in navigating) actually fired in constant alternation at 8 Hz as rats weighed alternative routes to take in a maze. This likely kept the images constantly alive in working memory. This rhythmic firing of the different cells could enable the alternatives to remain separate in imagination while uniting them in the same decision process. In this way, oscillating fields may serve as a basis of working memory and imagination. All these studies seriously challenge assumptions that neural EM fields play no role in brain circuits and cognition.

To summarize, these recent studies (and studies in previous sections) support our claims that qualia and their unity and causality arise from conscious EM activity. These studies challenge computationalist views such as IIT by indicating that *qualia aren’t seated in abstract, integrated information—they’re seated in the unified substance of EM fields and charges*. Simply put, it’s hard for IIT to explain how neural computations encode (for example) colors and their unified, pictorial forms—and how these codes yield actual conscious images. But our field theories offer evidence that conscious colors are simply laid out neuroelectrically in unified, pictorial form within neural maps (beyond how they appear to EEGs and other instruments).

Yet our MFT and DFT align with IIT in some ways. In our theories, thalamocortical feedback loops create fully unified consciousness because their synchrony helps create powerful EM fields. We thus end up ascribing consciousness largely to integrated, differentiated activity from reverberating brain areas, as IIT does. (IIT’s PCI of 0.31 may simply reflect the strong electrical activity and fully unified consciousness of these loops.) Also, our electrical approach makes similar predictions to IIT concerning split brains and cerebellums (§1.3). These structures lack the systematic electrical connections that local currents and fields use to unify experience.

**5. Conclusions**

Computationalist claims that minds are computational systems gain support from the highly computational structure of many neural operations. Nonetheless, conscious sensory and emotional qualities (qualia) may be noncomputational. For example, sensory-detector circuits process their signals too similarly to encode all our starkly different qualia computationally. Nor do these circuits bind their qualia (e.g., colored shapes) together to form overall unified images by means of any known code (synaptic, synchronic, etc.). Nor do computationalists explain in clear, intelligible ways how the qualia we feel are related to abstract (multiply realizable) codes.

“Computations” do undoubtedly modulate sensory circuits behind the scenes. But there’s mounting evidence that qualia themselves are concrete substances—not abstract computations—in these circuits. For example, different sensory qualia evidently correlate with different sensory detectors whose activities are EM energy substances. Also, qualia evidently bind together to form unified percepts by means of the detectors’ electrical activity, which is a continuous, unified EM-field substance. Reductionism is avoided here by treating these perceptions as the real, underlying substances of electrochemical activity in sensory maps—beyond how it appears to instruments such as EEGs. Reductionism is also avoided because the conscious neural EM field has irreducible dynamics and evidently interacts with circuits to help guide brain activity.

**References**

Aaronson, S. (2014) *Why I Am Not An Integrated Information Theorist* Accessed December 1

2014. http://www.scottaaronson.com/blog/?p=1799]

Aizawa, K. and Gillett, C. (2009) Levels, Individual Variation, and Massive Multiple

 Realization in Neurobiology, in Bickle, J. (ed.) Oxford Handbook of Philosophy and

 Neuroscience, New York: Oxford University Press.

Baars, B. (1988) *A Cognitive Theory of Consciousness.* Cambridge University Press.

Bahramisharif, A., Jensen, O., Jacobs, J., Lisman. J. (2018) Serial representation of items during

 working memory maintenance at letter-selective cortical sites. PLoS Biol 16 (8).

Bastos, A.; Loonisa, R.; Kornblitha, S.; Lundqvista, M. Miller, E. (2018) Laminar recordings in

frontal cortex suggest distinct layers for maintenance and control of working memory.

*PNAS* (115) 5:1117–1122.

Bayerl, D. and Bosch, O. (2019) Brain vasopressin signaling modulates aspects of maternal

 behavior in lactating rats. *Genes Brain Behav*. 2019Jan;18(1):e12517.

Berendzen, K. et al. (2023) Oxytocin receptor is not required for social attachment in prairie

 voles. Neuron. doi.org/10.1016/j.neuron.2022.12.011.

Bunge, M. (1979) Ontology, ii: *A World of Systems*.D. Reidel (Dordrecht).

Busnelli M., Chini B. (2017) Molecular Basis of Oxytocin Receptor Signalling in the Brain:

What We Know and What We Need to Know. In: Hurlemann R., Grinevich V. (eds)

Behavioral Pharmacology of Neuropeptides: Oxytocin. *Current Topics in Behavioral*

*Neurosciences*, vol 35

Cerullo, M. (2015). The problem with phi: A critique of integrated information theory. *PLoS*

*Computational Biology* 11(9), e1004286.]

Chalmers, D. 1996. *The Conscious Mind.* Oxford: Oxford University Press.

Chiang, C., R. Shivacharan, X. Wei, L. Gonzalez–Reyes, D. Durand. (2019) Slow periodic

activity in the longitudinal hippocampal slice can self–propagate non-synaptically by a

mechanism consistent with ephaptic coupling. *J Physiol* (597).

Churchland, Patricia (1986) *Neurophilosophy*, Cambridge, MA: MIT Press.

Churchland, Patricia (2013) *Touching a Nerve* Norton (NY, London).

Churchland, Paul (1981) Eliminative Materialism and the Propositional Attitudes, *Journal of*

*Philosophy* 78:2.

Churchland, Paul (1986a) A Neurocomputational Perspective: The Nature of Mind and the

 Structure of Science. MIT Press, Cambridge, Massachusetts.

Churchland, Paul (1986b) Some Reductive Strategies in Cognitive Neurobiology, *Mind* 95:379.

Conway, B. (2009) Color vision, cones, and color-coding in the cortex. Review Neuroscientist

 15(3):274-90.

Crick, F. & Koch, C. (2003) *A Framework for Consciousness*. *Neuroscience* vol. 6, 119–26.

Crick and Koch Toward a neurobiological theory of consciousness *Francis Crick & Christof*

 *Koch Seminars in the Neurosciences* 2:263-275 (1990)

Dehaene, S. (2014). *Consciousness and the brain: Deciphering how the brain codes our*

*thoughts.* Viking (New York).

Dehaene, S.; Naccache, L. (2001) Towards a cognitive neuroscience of consciousness: basic

evidence and a workspace framework. *Cognition* 79 1-37.

Dehaene S., Changeux J., Naccache L. (2011) The Global Neuronal Workspace Model of

Conscious Access: From Neuronal Architectures to Clinical Applications. In: Dehaene

S., Christen Y. (eds) *Characterizing Consciousness: From Cognition to the Clinic. Research and Perspectives in Neurosciences*. Springer (Berlin),

Ewing, A. (1962) *The Fundamental Questions of Philosophy.* Collier Books (New York).

Ezzyat et al., 2018 Ezzyat, Y. et al.—23 authors (2018) Closed-loop stimulation of temporal

cortex rescues functional networks and improves memory. *Nature Communications* 9 (1).

Fisher, H. (1997) Lust, attraction, and attachment in mammalian reproduction. *Human Nature*

9(1)23-52.

Fodor, J., 1975, *The Language of Thought,* New York: Thomas Y. Crowell.

Gourdie, R. (2019) The Cardiac Gap Junction has Discrete Functions in Electrotonic and

 Ephaptic Coupling, The Anatomical Record, 302, pp. 93–100.

Gray, C. & Singer, W. (1989) Stimulus specific neuronal oscillations in orientation columns of

 cat visual cortex. *Proc. Nati. Acad. Sci. USA* 86:1698 1702.

Haugeland, J. (1985) *Artificial Intelligence: The Very Idea.* MIT Press (Cambridge).

Heusser, A. Poeppel, D., Ezzyat, Y., Davachi, L. (2016) Episodic sequence memory is supported

 by a theta-gamma phase code, Nat Neurosci., 19 (10): 1374–1380.

Hales, C. (2014), “The origins of the brain’s endogenous electromagnetic field and its

 relationship to provision of consciousness,” Journal of Integrative Neuroscience 13.

Hales, C. & Ericson, M. (2022) Electromagnetism’s Bridge Across the Explanatory Gap: How a

 Neuroscience/Physics Collaboration Delivers Explanation Into All Theories of

 Consciousness. Front. Hum. Neurosci. Volume 16.

Hunt, T. and Schooler, J. (2019) The Easy Part of the Hard Problem: A Resonance Theory of

 Consciousness. *Frontiers in Human Neuroscience*, 13, Article No. 378.

Hunt, T. (2020) Calculating the Boundaries of Consciousness in General Resonance Theory.

 Journal of Consciousness Studies, 27, No. 11–12, 2020

Hunt, T., Ericson, M., & Schooler, J. (2022). Where’s My Consciousness-Ometer? How to Test

 for the Presence and Complexity of Consciousness. Perspectives on Psychological

 Science, 17(4), 1150–1165. https://doi.org/10.1177/17456916211029942

Jones, M. (2013) Electromagnetic-Field Theories of Mind, Journal of Consciousness Studies 20

 (11-12): 124-149.

Jones, M., (2016a) Avoiding Perennial Mind-Body Problems, Journal of Consciousness Studies

 23 (9-10): 111-133.

Jones, M. (2016b) Neuroelectrical approaches to binding problems, Journal of Mind and

 Behavior 2 (37).

Jones, M. (2017) Mounting Evidence that Minds Are Neural EM Fields Interacting with Brains,

 Journal of Consciousness Studies 24 (1-2): 159-183.

Jones, M. (2019) Growing Evidence that Perceptual Qualia are Neuroelectrical Not

 Computational, Journal of Consciousness Studies, 26 (5-6): 89-116.

Jones, M.; LaRock, E. (2019) How Subjects Can Emerge from Neurons, Process Studies, 48 (1):

 40-58.

Jones, M.; LaRock, E. (2021) From Murphy's Christian Physicalism to Lowe's Dualism. In

 Mihretu Guta, Ph.D., and Eric LaRock, Ph.D., (editors), E. J. Lowe's Metaphysics and

 Analytic Theology, Theologica: An International Journal for Philosophy of Religion and

 Philosophical Theology 5 (2) 100-128: (UC Louvain).

K. E. K. Hering (1913) Scientific papers, published by the Saxon Academy of Sciences. Leipzig

1913.

Kay, K., Chung, J., Sosa, M., Schor, J., Karlsson, M., Larkin, M., Liu, D., Frank, L. (2020)

Constant Sub-second Cycling between Representations of Possible Futures in the

Hippocampus. *Cell* 180 (3): 552-567.

Koch C.; Massimini, M; Boly, M.; Tononi, G. (2016) Neural correlates of consciousness:

 progress and problems. *Nature Review Neuroscience* 17.

Koch (2019) *The Feeling of Life Itself—Why Consciousness is Widespread but Can’t be*

*Computed*, MIT Press (Cambridge).

Kosslyn, S. (1983) *Ghosts in the Mind’s Machine*. New York: Norton and Company.

LaRock, E., (2007) Disambiguation, Binding, and the Unity of Visual Consciousness. Theory &

 Psychology, vol. 17, pp. 747-777.

Larock, E. (2019) Hard Problems of Unified Experience from the Perspective of Neuroscience.

 In M. Guta, Ph.D., (editor), Consciousness and the Ontology of Properties. New York:

 Routledge.

Larock, E.; Schwartz, J.; Ivanov, I.; Carreon, D. (2020) A Strong Emergence Hypothesis of

 Conscious Integration and Neural Rewiring. International Philosophical Quarterly, Vol.

 60, pp. 97-115.

Leibniz (1695) New System of Nature. In: *Leibniz Philosophical Writings*, G Parkinson (ed).

Dent & Sons 1973 (London).

Leibniz (1714) Monadology. *In: Leibniz Philosophical Writings*, G Parkinson (ed). Dent & Sons

 1973 (London).

Levine, J. (1993) On Leaving Out What It’s Like. In: M. Davies and G. Humphreys (eds)

*Consciousness: Psychological and Philosophical Essays*. Oxford: Blackwell.

Libet, B. (1993) *Neurophysiology of Consciousness*, Boston, MA: Birkhauser.

Lindahl, B.; Arhem, P. (1994) Mind as a force field. *Journal of Theoretical Biology*

171:111-122.

Lisman, J. and Jensen, O. (2013) The Theta-Gamma Neural Code, Neuron, 77 (6) pp. 1002

 -1016.

Liu, H. (2019) Of integrated information theory: a philosophical evaluation, *Philosophical*

*Psychology* DOI: 10.1080/09515089.2019.1695044

Lorenz, K. (1966) *On Aggression* Methuen.

Lorenz, K. (1977) *Behind the Mirror.* Harcourt Brace Jovanovich.

Lovheim, H. (2012) A new three-dimensional model for emotions and monoamine

Neurotransmitters. *Med Hypotheses* 78(2):341-8.

Lowe, E. J. (2010) *Substance Dualism: A Non-Cartesian Approach.* In Koons, R. & Bealer G.

(eds) *The Waning of Materialism.* Oxford University Press (Oxford)

Martinez-Banaclocha, M. (2020) Astroglial Isopotentiality and Calcium-Associated Biomagnetic

 Field Effects on Cortical Neuronal Coupling, Cells, 9, 439.

McCulloch, W. & W. Pitts (1943) A Logical Calculus of the Ideas Immanent in Nervous

Activity. *Bulletin of Mathematical Biophysics* 7: 115-133.

McFadden, J. (2013), The CEMI field theory: closing the loop, *Journal of Consciousness*

 *Studies*, 20:1-2;153-168(16).

Meyer, M., Corcoran,K. Chen, J., Gallego, S., Li, G., Tiruveedhula, V., Cook, J., Radulovic, J.

 (2017) Neurobiological correlates of state-dependent context fear Cold Spring Harbor

 Laboratory Press 24: 385–391.

Milkowski, M. (2018) Objections to computationalism: a survey. Roczniki Filozoficzne. Tom

LXVI, number 3.

Mindt, G. (2017) The problem with the ‘Information’ in integrated information theory. *Journal*

*of Consciousness Studies* 24(7–8): 130–154.

Murphy, N. & Brown, W. (2007) *Did My Neurons Make Me Do It?* Oxford University Press).

Narikiyo, K., Mizuguchi, R., Ajima, A., Shiozaki, M., Hamanaka, H. Johansen, J., Mori, K.,

 Yoshihara, Y. (2020) The claustrum coordinates cortical slow-wave activity nature

 neuroscience, Nat Neurosci. 23 (6) pp. 741-753.

Newell, A. & Simon, H. (1956) The Logic Theory Machine: A Complex Information Processing

System. *IRE Transactions on Information Theory*, IT-2, 3: 61-79.

Parker, D. (2019). Color Perception. SAGE Publications. Young, T. (1802). II. The Bakerian

 Lecture. On the theory of light and colors. Philosophical Transactions of the Royal

 Society of London, (92), 12-48.

Piccinini, G. & Scarantino, A. (1986) Information processing, computation, and cognition. J Biol

 Phys. 37(1).

Pockett, S. (2006), “The Neuroscience of movement”, *Does Consciousness Cause Behavior?*

 (ed) S. Pockett, W. Banks, S. Gallagher (Cambridge: MIT Press).

Ptito M.; Schneider F.; Paulson O.; Kupers R. (2008) Alterations of the visual pathways in

congenital blindness. *Exp. Brain Res*. 87: 41–4910.

Purves D, Augustine GJ, Fitzpatrick D, et al., (2001) Neural Coding in the Taste System, in

 Neuroscience. 2nd edition, Sunderland (MA): Sinauer Associates.

Putnam, H. (1967) The Nature of Mental States, in Capitan, W. & Merrill, D. (eds)

*Art, Mind, And Religion*. University of Pittsburgh Press.

Rescorla, M. (2015) The Computational Theory of Mind. In: Zalta, E. (ed.) The Stanford

Encyclopedia of Philosophy.

Riddle, J., Scimeca, J., Cellier, D., Dhanani, S., D’Esposito, M. (2020) Causal Evidence for a

 Role of Theta and Alpha Oscillations in the Control of Working Memory. Current

 Biology 30 pp. 1748–1754.

Rosenberg, G. (2004), *A Place for Consciousness.* Oxford University Press.

Ross, H. & Young, L. (2009) Oxytocin and the Neural Mechanisms Regulating Social Cognition

and Affiliative Behavior. *Front Neuroendocrinol,* 30(4): 534–547.

Rumelhart, D.; McClelland, J.; and the PDP Research Group, 1986, *Parallel Distributed*

*Processing*, vol. 1. Cambridge: MIT Press.

Russell, B. (1927/1954), The Analysis of Matter (New York: Dover).

Schafer, M. and Schiller, D. (2018) Navigating Social Space. Neuron, 100 (2) pp. 476–489.

Shaffer, J. (1968) *Philosophy of Mind*, Engle Woods Cliff NJ: Prentice-Hall.

Shivacharan, R; Chiang, C.; Zhang, M; Gonzalez-Reyes, L; Durand, D. (2019) Self-propagating,

non-synaptic epileptiform activity recruits neurons by endogenous electric ﬁelds,

*Experimental Neurology* 317:119–128.

Solomon SG, Lennie P. (2007) The machinery of colour vision. Nat Rev Neurosci. 8(4):276-86.

Sperry, R. (1952) Neurology and the mind–brain problem, *American Scientist*: 291–312.

Sprouse-Blum, A.; Smith, G.; Sugai, D.; Parsa, D. (2010) Understanding Endorphins and Their

 Importance in Pain Management. Hawaii Med J. 69(3):70–71.

Strawson, G. (2016) Mind and being: the primacy of panpsychism. In: Bruntrup, G. & Jaskolla,

L. (eds), *Panpsychism: Philosophical Essays*. Oxford University Press.

Su A, et al. (2004) A gene atlas of the mouse and human protein-encoding transcriptomes.

*Proc. Natl. Acad. Sci. U.S.A.* 101(16): 6062–7.

Turing. A. (1936) On computable numbers, with an application to the entscheidungs problem.

Proceedings of the London Mathematical Society 42: 230-265.

Tokera, D.; Pappasb, I.; Lendnerb, J.; Frohlicha, J.; Mateosf, D.; Muthukumaraswamyi, S.;

 Carhart-Harrisj, R.; Paffl, M.; Vespam, P.; Montia, M.; Sommerb, F.; Knightb, R.;

 D’Esposito, M. (2022) Consciousness is supported by near-critical slow cortical

 electrodynamics, PNAS 119:7

Tononi, G. (2008) Consciousness as Integrated Information: a Provisional Manifesto. *Biol. Bull.*

215:216-42.

Tononi, G. (2012) Integrated information theory of consciousness: an updated account. *Archives*

*Italiennes de Biologie.* 150: 290-326.

Tononi, G. & Koch, C. (2008) Consciousness as Integrated Information: a Provisional

Manifesto. *Biol. Bull.* 215: 216 –242.

Tononi, G. & Koch, C. (2015) Consciousness: here, there and everywhere? Philos Trans R Soc

 Lond B Biol Sci . 370(1668):20140167.

Tononi, G.; Boly, M.; Massimini, M.; Koch, C. (2016) Integrated information theory: from

 consciousness to its physical substrate. Nature Reviews: Neuroscience, vol. 17

Treisman, A. (2003) Consciousness and Perceptual Binding. In: *The Unity of Consciousness,*

A. Cleeremans (ed). Oxford University Press.

Tsuchiya, N. & Koch, C. (2008) Attention and Consciousness. *Scholarpedia*, 3(5):4173.

Vernet, M., Stengel, C, Quentin, R., Amengual, J. Valero-Cabré, A. (2019) Entrainment of local

 synchrony reveals a causal role for high-beta right frontal oscillations in human visual

 consciousness, Nature: Scientific Reports 9.

Zeki, S. (1993) *A Vision of the Brain*. Blackwell (London).

Zeki, S. (2003) The Disunity of Consciousness. *Trends in Cognitive Sciences* 7:214–1050 218.

Zeki, S. (2015. A Massively Asynchronous, Parallel Brain. *Phil Trans. R Soc* B370:1–1052 14.

Zhang, Y. Tsang, T., Bushong, E., Chu, L., Chiang, A., Ellisman, M., Reingruber, J. Su, C.,

 (2019) Asymmetric ephaptic inhibition between compartmentalized olfactory receptor

 neurons, Nature Communications, 10 (1), 1560.

1. Computationalism might be instead construed as the incontestable claim that that minds merely require computations (Milkowski, 2018). However, this paper is concerned with the contestable issue of whether minds can be explained in purely computational terms, as computationalism usually contends. [↑](#footnote-ref-1)
2. In nonreductive physicalism, mental events aren’t reducible to physical events, but still depend (supervene) on them. Putnam’s functionalist idea of multiple realization clarifies nonreductive physicalism’s idea of supervenience by explaining how minds can depend on physical events by being functional states realized in certain physical states. [↑](#footnote-ref-2)
3. Computations are often associated more with classical approaches, while information processing is often associated with connectionist approaches (Piccinini & Scarantino, 1986), but for our purposes computations and information processing needn’t be distinguished. [↑](#footnote-ref-3)
4. Hunt investigates better measures of consciousness to assess IIT and related theories such as his General Resonance Theory, which links consciousness instead to shared resonances within fields and associated phase transitions to richer, more complex consciousness. [↑](#footnote-ref-4)
5. Russell said here that “Physics, in itself, is exceedingly abstract, and reveals only certain mathematical characteristics” of matter, not its “intrinsic” character (1927, p. 10). [↑](#footnote-ref-5)
6. This helps IIT reply to Mindt’s (2017) claim that IIT runs afoul of Chalmers’ (2003) argument that explanations based on the structure and dynamics of physical processes yield just more structure and dynamics, so structures and dynamics are all we can expect such processes to explain. IIT arguably escapes this criticism by treating qualia as the intrinsic nature of these structures and dynamics. Yet IIT (like other Russellian monisms) is still obscure about how qualia can be the intrinsic nature of abstract relations. Also, IIT may be open to Aaronson’s (2014) argument that IIT implies something quite dubious: that even a simple DVD player’s integrated information can be conscious. [↑](#footnote-ref-6)
7. For example, Lovheim offers no better guidance than Churchland on where the emotion of love comes from. He presumably puts it in the emotion of joy. But the joy of romantic love differs greatly from the joy of monetary riches. Arguably, his preoccupation with the “basic” emotions such as excitement, joy, and surprise misleads him into treating joy as a single emotion with a single cause. This neglects the rich variety of emotions and hormones. [↑](#footnote-ref-7)
8. Hales & Ericson (2022) argue that “the brain is an essentially electromagnetic (EM) field object from the atomic level up,” in that EM “has primacy” in establishing consciousness and is the basis of everything relevant to life. This aligns with our view that qualia arise in EM fields and EM charges (especially if we follow some physicists in treating charged particles as fields). [↑](#footnote-ref-8)
9. We refer to this theory as monist instead of materialist because materialism (and physicalism) are so often associated with ontological and causal reductionism. [↑](#footnote-ref-9)
10. Lowe treats substances as fundamental bearers of properties, while we treat substances in a loosely related way as the fundamental kinds of stuff comprising the universe. [↑](#footnote-ref-10)
11. DFT differs from current dualisms in which consciousness emerges from nonconscious complexity. The latter is widely criticized for being akin to magic. By contrast, in DFT, consciousness is fundamental (it’s EM energy). It’s not emergent from complexity. However, the conscious subject does emerge from simpler conscious components. [↑](#footnote-ref-11)
12. Yet subjects can possess nonconscious states, and undergo unconscious states, subliminal priming, et cetera. [↑](#footnote-ref-12)
13. IIT avoids this particular problem, but its abstract, complex, untestable approach still renders qualia quite elusive. [↑](#footnote-ref-13)
14. Our claim is that each protein has a different quality. This is compatible with different proteins having the same quality (see Jones, 2019). [↑](#footnote-ref-14)
15. It’s possible that qualia may be affected not only by channels and GPCRs but also by nearby active components they strongly interact with (see Jones, 2019). However, the details are quite unclear, so this paper sticks to what seems relatively clear at present—the link between qualia and channels/GPCRs. Once these interactions become more clear, they’ll be incorporated into our two theories. [↑](#footnote-ref-15)
16. What exactly is it about these proteins that gives them their different qualia? It can’t be their different molecular structures, for these are emergent, while MFT/DFT treat consciousness as a fundamental substance. Instead, qualia could be the different resonances of the proteins or their fields. These have fundamental energy (frequency) levels. But it’s not yet known which resonant energies would correlate with qualia. Alternatively, Jones (2019) shows that each electrically active detector protein correlates with a distinct mass. This suggests that qualia are ultimately tied to narrow bands of electrically bound masses (and thus rest energies). In fact, the entire range of our qualia may reside like a rainbow in the range of these protein masses. In nature at large, this rainbow would likely repeat across many orders of magnitudes of electrically bound masses. [↑](#footnote-ref-16)
17. In MFT, the qualia in our visual images come from electrically active membrane proteins, while the neural field serves to unify these colors to form the images. The field is too insubstantial relative to the proteins (see footnote 16 on rest energies) to determine the colors in images, but the field’s continuity across space is well suited to unifying colors. Also, by unifying mental operations into a single whole, the field becomes the vital arena for short-term memory and imagination (see §4.4). [↑](#footnote-ref-17)
18. One problem for claims that synchronized firing binds diverse brain brain activities into a conscious, unified whole is that this synchrony occurs during seizures which lack consciousness. Binding by coherent fields avoids this problem by explaining why seizures aren’t conscious. These electric storms disrupt the oscillating electrical fields localized along circuits that run between brain areas. Such circuits normally help to carry out information processing between areas and to sculpt the focus of attention, select the content of working memory, and conduct the operations of imagination and other cognitive activities (Bastos et al., 2018; Ezzyat et al., 2018; Kay et al., 2020). When the electrical activity of circuits is upset by seizures, attentive coordinations between brain areas break down. The EM fields that unify consciousness are disrupted (Huang et al., 2018). In a similar context, Toker et al. (2021) argue that “the electric activity of the cortex is indeed poised near the boundary between stability and chaos during conscious states and transitions away from this boundary during unconsciousness . . . disrupts cortical information processing.” [↑](#footnote-ref-18)
19. For example, Hunt (2022) argues that “parts of a whole resonate at the same or similar frequencies. These parts, in any complex consciousness such as exists in animals, exist in a nested hierarchy of faster frequencies at more fundamental levels in the hierarchy to slower frequencies at the top. The phase transitions in information transfer made possible through such shared resonance is the key mechanism that allows for the unity of consciousness in each moment.” [↑](#footnote-ref-19)
20. This explanation of mental unity in our theories forms the basis of an overall account of how we avoid panpsychism’s so-called combination problem. This problem—which we deal with in Jones, 2016b—concerns how conscious neurons, molecules, etc. combine to form our overall perceptions, memories, and the overall subject. [↑](#footnote-ref-20)