

Rejecting the "formless stuff" response to the origin of consciousness problem for IIT 4.0

Refutación de la respuesta de la "materia informe" al problema del origen de la conciencia en la IIT 4.0

Ignacio Cea^{1,2}✉

Abstract

This paper critically evaluates and ultimately rejects a proposed solution to a problem arising from the ontological commitments of Integrated Information Theory (IIT) 4.0, particularly its assertion that only conscious entities truly exist. This ontological commitment generates the "origin of consciousness problem": if non-conscious entities are denied genuine existence, it becomes unclear how consciousness could evolve from non-conscious ancestors. As Cea et al. (2024a) argue, this issue leads IIT toward a problematic form of "Big Bang consciousness," a primordial ontological dust of minimally conscious elementary particles originating at the beginning of the universe. This paper analyzes and ultimately rejects an alternative solution to the origin of consciousness problem, the "formless stuff hypothesis" (Koch, 2024). This view seeks to avoid both the notion that consciousness arises from nothing and the requirement for "Big Bang consciousness." Instead, it posits that, prior to conscious experience, a formless, undifferentiated state of reality exists, distinct from nothingness. However, this proposal is implausible for three reasons: i) it faces challenges internal to IIT, ii) it conflicts with current scientific understanding of the non-conscious world, and iii) Koch's primary justification for introducing 'formless stuff' is based on an epistemological-ontological conflation. It is therefore recommended that IIT adopt a form of causal-physical realism, which maintains that a structured and lawful universe exists independent of consciousness.

Keywords: *Metaphysics of consciousness, Evolution of consciousness, Ontological dust, Cosmology of consciousness, Great Divide of Being*

Resumen

Este artículo examina y finalmente rechaza una solución propuesta a un problema derivado de los compromisos ontológicos de la Teoría de la Información Integrada (IIT) 4.0, en particular su afirmación de que solo las entidades conscientes existen verdaderamente. Este compromiso genera el "problema del origen de la conciencia": si se niega la existencia genuina de las entidades no conscientes, ¿cómo podría la conciencia haber evolucionado a partir de ancestros no conscientes? Como argumentan Cea et al. (2024a), este problema fuerza a la IIT hacia una forma altamente problemática de "conciencia Big-Bang": un polvo ontológico primordial de partículas elementales mínimamente conscientes originadas al comienzo del universo. Aquí analizamos y finalmente rechazamos una solución alternativa al problema del origen de la conciencia basada en lo que llamamos la "hipótesis de la materia informe" (Koch, 2024). Esta propuesta intenta evitar tanto la idea de que la conciencia surge de la nada como la necesidad de una "conciencia Big-Bang", postulando en cambio que, antes de la experiencia consciente, existe un estado de realidad informe e indiferenciado, pero distinto de la nada. Sin embargo, sostenemos que esta propuesta es implausible i) por razones internas a IIT, ii) por consideraciones relacionadas con nuestro conocimiento científico del mundo no consciente, y iii) porque la motivación principal de Koch para introducir esta 'materia informe' descansa en una confusión epistemológico-ontológica. En cambio, IIT debería adoptar una forma de realismo causal-físico según el cual existe verdaderamente un universo estructurado y regido por leyes más allá de la conciencia.

Palabras clave: *Metafísica de la conciencia, Evolución de la conciencia, Polvo ontológico, Cosmología de la conciencia, Gran División del Ser*

1 Facultad de Ciencias Religiosas y Filosofía, Universidad Católica de Temuco, Chile, 2 Facultad de Filosofía y Humanidades, Universidad Alberto Hurtado Chile ✉ e-mail: igneocj@gmail.com



Introduction: IIT 4.0 essentials

Integrated Information Theory (IIT) is currently one of the most debated scientific approaches to understanding consciousness (Cleeremans et al., 2025; Consortium et al., 2025; Seth & Bayne, 2022; Signorelli et al., 2021; Storm et al., 2024). Unlike other well-known theories, such as the Global Neuronal Workspace Theory (Dehaene, 2014; Dehaene & Naccache, 2001; Mashour et al., 2020), or the Attention Schema Theory (Graziano, 2017, 2019), the theoretical starting point of IIT is a description of consciousness itself, as experienced from the first-person perspective, rather than its neural, behavioral, computational, or functional correlates (Chis-Ciure, 2022; Ellia et al., 2021; Negro, 2020, 2022a).

More specifically, IIT is constructed from a set of first principles known as the "axioms of phenomenal existence" (Albantakis et al., 2023; Comolatti et al., 2025; Tononi et al., 2022; Tononi & Boly, 2025). The most basic of these is the undeniable and directly known fact that consciousness exists, referred to as the "zeroth axiom". Building on this foundation, IIT identifies five key properties that are claimed to be essential to consciousness: intrinsicity, information, integration, exclusion, and composition.

These properties are then converted into scientifically applicable constructs, known as the "postulates of physical existence". These postulates transform the essential properties of experience into measurable, operational characteristics that define what it means for a physical system to be conscious. The theory mathematically formalizes these postulates, identifying the "complex", also called the physical substrate of consciousness, as the system that achieves the *maximal system integrated information* (ϕ_s^*), while "overlapping substrates with lower ϕ s are *excluded from existence*" (Albantakis et al., 2023, p. 12, italics added)¹.

Within Integrated Information Theory (IIT), the concept of "exclusion from existence" is interpreted literally rather than metaphorically. IIT asserts that candidate substrates failing to specify the maximal value of system integrated information (ϕ_s^*) do not truly exist, as they lack existence "for themselves" as conscious entities. According to IIT, only conscious experiences truly exist because only they "exist for themselves", in an absolute manner, only the existence of conscious entities is immediately and irrefutably known by the entities themselves. Therefore, only complexes with maximal ϕ_s truly exist, as they are the ones that exist consciously and thus "for themselves" in the relevant phenomenal sense. In short, entities that truly exist in objective reality are limited to those that intrinsically exist as subjective experiences (Cea et al., 2023)². These entities exist in themselves, independently of external observers.

This idea provides the foundation for IIT's "Great Divide of Being" (Koch, 2024; Tononi, 2017; Tononi et al., 2022; Tononi & Boly, 2025), which refers to "the divide between what truly exists in an absolute sense, in and of itself —namely conscious, intrinsic entities— and what only exist in a relative sense, for something else" (Tononi et al., 2022, p. 8). In other words, only conscious entities that exist intrinsically truly exist, whereas non-conscious extrinsic entities "exist", at most, only from the perspective of, or *for*, another intrinsic entity. According to the theory, "extrinsic existence is grounded in

¹ In the IIT 4.0 formalism, there are other important measures like the *integrated information for distinctions* $\phi_{\mathcal{P}}$, *integrated information for relations* ϕ_r , *structure integrated information* Φ , and the Φ -structure. However, given the scope of this article, the focus is on the *system integrated information* ϕ_s , and in particular, its maximal value among overlapping systems (ϕ_s^*). The reader is referred to Albantakis et al. (2023) for details about the other measures and the whole IIT 4.0 formalism.

² However, as discussed below, this leaves open the possibility that nonconscious objects may still exist as reducible aggregates of truly existing, minimally conscious parts.

intrinsic existence... extrinsic interactions can be considered properties of intrinsic existence" (Grasso et al., 2024). In other words, extrinsic entities neither exist for themselves, independently of other entities, nor are nothing, but exist in virtue of being experienced by another intrinsic entity:

Bodies and organs, tables and rocks, stars and planets. . . are likely to unfold into extrinsic entities. . . They only exist vicariously, from the perspective of some intrinsic entity, and so they do not truly exist. (Tononi et al., 2022, p. 8)

That is, IIT adheres to what Cea and colleagues (2023) call IIT's *principle of true existence*, namely that "only phenomenal existence is true existence" (p. 4), whereas familiar large-scale objects that presumably do not maximize ϕ_s , like bodies and stars, at best qualify as extrinsic entities that only exist mind-dependently, i.e. from the perspective of some other conscious, intrinsic entity that experiences them³. Thus, the existence of extrinsic entities may be compared, in a loose sense, to mind-dependent institutional facts such as money or governments (Searle, 1995): they are distinct from nothing, yet their existence is relative to the perspective of intrinsically existing systems. However, unlike Searle's institutional facts, which supervene on a mind-independent physical reality, extrinsic entities in IIT are grounded in intrinsic existence itself and thus do not presuppose a more fundamental observer-independent ontological base.

IIT's ontology is not solipsistic (i.e. only my consciousness exists), but rather a form of realism, meaning that IIT embraces the existence of an external world beyond one's own conscious experience (Albantakis et al., 2023; Koch, 2024; Tononi et al., 2022). However, IIT's realism is theoretically constrained by its own "Great Divide of Being" on the one hand, and the traditional meaning of being realist about "X" on the other hand, which normally entails: i) believing that "X" exists and ii) that "X"'s existence and properties are independent of my mind (Miller, 2021). Therefore, IIT's external-world realism, in the standard sense of the term, should be understood as claiming that an external world truly exists, beyond our own consciousness. However, IIT holds that this world is fundamentally constituted by other conscious, intrinsic entities, each one enjoying its own subjective experience. This stance can be understood as a form of *realist idealism* (Cea et al., 2023; Chalmers, 2021).

In the remaining sections, an important problem that follows from these ontological assumptions is presented: the "origin of consciousness problem". This is the question of how consciousness could have evolved from non-conscious ancestors if, according to IIT, without consciousness there is neither intrinsic nor extrinsic existence, i.e., the problem that conscious entities would have evolved from nothing (Cea et al., 2024a). Then, a potential solution to this problem based on Koch's "formless stuff hypothesis" (Koch, 2024) is presented and ultimately rejected. This potential solution suggests that before consciousness, there was something instead of nothing, a state of unstructured ontological chaos.

³ Note that the ϕ_s of systems such as human bodies or stars has not been measured, and doing so would involve serious technical challenges (e.g., specifying the relevant TPMs). However, our claim is conceptual and follows IIT theorists themselves, who state that "Bodies and organs..., stars and planets... are likely to unfold into extrinsic entities" (Tononi et al., 2022, p. 8). Thus, independently of whether such systems eventually turn out to maximize ϕ_s or not, IIT implies that only ϕ_s -maximizing systems qualify as intrinsic entities that truly exist, while all others count as extrinsic and mind-dependent.

The "origin of consciousness problem" for IIT.

In the science of consciousness, it is widely accepted that human and non-human animal consciousness evolved from earlier forms of experience possessed by our evolutionary ancestors, which themselves evolved from even more primitive forms, tracing back the origins of sentience to some point(s) in the evolutionary tree of life (Feinberg & Mallatt, 2013, 2016; Ginsburg & Jablonka, 2019; Godfrey-Smith, 2016, 2024; Veit, 2023). In other words, non-conscious forms of life eventually evolved into primordial forms of consciousness, probably due to the emergence of the right sort of nervous systems in both vertebrate and invertebrate animals (Birch, 2024; Damasio & Damasio, 2024; Denton, 2005; Humphrey, 2022).

It is an open debate whether this evolutionary transition from non-conscious to conscious organisms had a sharp boundary from total absence of consciousness to its minimal form, or a gradual transition with indeterminate borderline cases (Birch, 2024; Godfrey-Smith, 2020). However, irrespective of this controversy, most people agree with what can be called "the received view": conscious creatures phylogenetically evolved from non-conscious organisms. The received view is nicely captured by Humphrey:

Today, sentience may--possibly--be all around us. But there was a time in the history of living organisms when it did not exist anywhere on Earth. Given that humans had non-sentient ancestors, there has to be a story to be told about how our ancestors got from there to here. (Humphrey 2022, p. 101)

In IIT's terms, this means that natural selection may have pressed organisms to evolve their fully modular nervous systems into progressively more strongly integrated and recursive networks with increasingly higher integrated information (Albantakis et al., 2014)⁴. Indeed, according to Koch (2024), "the first flickering of phenomenal light likely appeared during the Cambrian explosion, 530 million years ago, as multicellular animals and their primitive nervous nets arose and proliferated" (p. 105). Thus, IIT seems able to account for the evolution of consciousness within a biological context (see also Tononi & Koch, (2015)), even if its theoretical constructs are not necessarily associated with life-sustaining processes and are compatible with non-biological consciousness (Ellia & Chis-Ciure, 2022; Koch, 2024).

However, the problem is that IIT's current ontology entails that ancient non-conscious organisms did not *truly* exist, precisely because they had no consciousness. At best, they could have only existed from the vicarious perspective of other conscious (intrinsic) entities. But if the view under consideration is correct, and there was indeed a period in Earth's history during which consciousness was entirely absent, then our non-conscious ancestors could not have existed extrinsically either, since there was no intrinsic entity for which they could exist. Thus, if the received view is right and consciousness first evolved from previously non-conscious entities, IIT entails that consciousness originated from non-existent stuff, i.e., from nothing. This is what is called the "origin of consciousness problem" for IIT: how could consciousness have evolved from non-conscious mechanisms in ancestors that lacked sentience, if, according to IIT, any non-conscious entity prior to the origin of consciousness should be considered non-existent (i.e., neither an intrinsic nor an extrinsic entity)? In other words, if the received view is correct, then IIT entails that consciousness phylogenetically evolved from nothing.

⁴ It remains an open and largely speculative question whether early nervous systems were fully modular, or whether some degree of integration was already present in the earliest phylogenetic stages. In particular, some examples discussed in the IIT literature rely on simplified or artificial models rather than empirical connectomic data.

It has been argued that to account for the origin of consciousness, IIT is committed to posit a *primordial ontological dust*: a collection of intrinsically existing, minimally conscious, elementary particles without internal structure that emerged at the beginning of the universe, i.e., a kind of "Big Bang (minimal) consciousness" (Cea et al., 2024a). However, this is deeply problematic for several reasons. First, it entails that indivisible monadic elements must specify positive values of system integrated information, which recent work has shown is mathematically, computationally, and conceptually inconsistent with IIT's own framework (Cea et al., 2025). Second, the notion of a primordial ontological dust is contrary to the plausible link between consciousness and life (Cea, 2023b, 2023a; Cea & Martínez-Pernía, 2023; Damasio & Damasio, 2024; Fuchs, 2018; Seth, 2021, 2024, 2025; Thompson, 2007, 2015), exemplified in the fact that all known conscious entities are living organisms (Birch, 2024; Godfrey-Smith, 2024; Thompson, 2007, 2015), which is also connected to the received view that consciousness probably evolved in previously non-conscious ancestors due to some significant adaptive advantage it provided (Denton et al., 2009; Ginsburg & Jablonka, 2019; Godfrey-Smith, 2024; Humphrey, 2022). Third, positing a primordial ontological dust conflicts with standard physical cosmology, which holds that the early universe's components and formative processes had absolutely nothing to do with consciousness (Peebles, 2020). In the next section, we present a potential way in which IIT might respond to the origin-of-consciousness problem without this problematic primordial ontological dust.

Koch's potential reply: the "formless stuff" hypothesis

If consciousness evolved from non-conscious processes, structures, or mechanisms, then IIT would be forced to claim that consciousness originated from literally nothing, because a universe without consciousness is a universe devoid of both intrinsic and extrinsic existence, which, according to IIT exhaust all that exists.

In fact, this is suggested by Tononi and colleagues (2022), when, quoting Schrodinger, they write that a universe without consciousness "would be 'a play before empty benches, not existing for anybody, thus quite properly speaking not existing'" (p. 8).

However, intriguingly, after repeating exactly the same quote from Schrodinger, Koch (2024) writes that "...Yes, without a conscious audience, there is no play. *Just stuff happening*" (p. 105, italics added). How, then, could there be "stuff happening" if there is neither intrinsic nor extrinsic existence, and both were meant to exhaust everything that exists? Koch also writes the following:

Think of the early Earth... four billion years ago, the planet was lifeless. Assuming that the integrated information of small assemblies of hydrocarbons is close to zero, ontological dust from the point of view of absolute existence, almost nothing existed for itself. *Despite the sun shining*, Earth's surface was dark, without the inner light of consciousness. (p. 105, italics added)

Again, it is asked, how could the Sun shine if there was, at best, just a tiny "light" of consciousness and intrinsic existence in the form of an ontological dust of minimally conscious hydrocarbons? The less problematic answer would be that the Sun shone, at best, only from the extrinsic perspective of these small assemblies of minimally conscious hydrocarbons. However, that does not seem to be what Koch means. He

suggests that before consciousness, there was *something* rather than *nothing* indeed. He writes:

This [IIT's Great Divide of Being] also answers the classic Philosophy 101 riddle: "If a tree falls in the forest, does it make a sound if no one is around to hear it?" Indeed, there is no sound without a conscious observer to hear. Furthermore, there isn't even a tree or a forest, as these concepts depend on a conscious subject discerning a tree from other trees, treating them as different from the soil they are planted in or from the air that surrounds them. Nature knows nothing of these distinctions, of trees and forests, but only of formless stuff. Without a conscious subject, there are only "atoms and the void," as Democritus stated, or *toho wa-bohu* (sic), as the Hebrew Bible describes the Earth before the act of creation in Genesis. (2024, p. 105)

Thus, Koch claims that without consciousness there is only a state of undifferentiated, formless, meaningless reality, not the structured, coherent, meaningful world that is experienced. Although Koch's notion of an unstructured state of existence in the absence of consciousness may be difficult to grasp, if plausible, it would certainly be better than *nothing*. Let us unpack Koch's suggestion in more detail.

Let us start with Koch's assertion that "without a conscious observer ... there isn't even a tree or a forest, as these concepts depend on a conscious subject discerning a tree from other trees... Nature knows nothing of these distinctions... but only of formless stuff" (2024, p. 105). According to IIT, conscious experience is a structure constituted by phenomenal distinctions (e.g., consciously perceived leaves, branches, trees, among others) that are interrelated in specific ways (e.g., a tree has branches with leaves, and is next to another tree with other branches and leaves). Thus, Koch claims that without consciousness, these interrelated phenomenal distinctions are absent, and hence the difference between a leaf and a branch, or a tree from another tree, would be inapplicable to nature as such, beyond our experience of it. Hence, without consciousness, there are no structured distinctions but just "formless stuff".

Interestingly, this view is akin to Kant's (1999), according to which the mind actively structures raw, formless sensory data through categories and concepts, thereby organizing experience into coherent objects like trees and forests. In Kant's philosophy, objects as they are known depend on the faculties of the human mind, implying that without a conscious subject, there is no phenomenon, only 'things-in-themselves' (i.e., noumena) that are inaccessible to us, much like Koch's 'formless stuff'.

Also, Koch claims that "without a conscious subject, there are only "atoms and the void" as Democritus stated, or *toho wa-bohu* (sic)" (2024, p. 105). *Tohu wa-Bohu* is meant to describe the chaotic, formless, and unstructured state of the Earth before the act of divine creation (Kister, 2007), while Democritus' "atoms and the void" arguably conveys the idea of a universe devoid of the meaning and experiential integration that consciousness supplies. Thus, Koch claims that without conscious experience, there is only Nature in a state of complete undifferentiation without form and meaning, something akin to an *ontological chaos*, but still different from *nothing*⁵.

Hence, Koch's potential answer to the origin-of-consciousness problem might be the following. First, recall that Koch (2024) adheres to what is called *the received view*: "the first flickering of phenomenal light likely appeared during the Cambrian explosion...as multicellular animals and their primitive nervous nets arose and proliferated" (p. 105). Thus, Koch may reply that consciousness evolved from non-conscious predecessors, but this does not imply that it originated from nothing. Consciousness could have emerged

⁵ "ontological chaos" is not meant as chaos in the technical sense of dynamical systems, nor a term explicitly used by Koch, but a heuristic characterization of a reality that, without intrinsic experience, would lack intrinsic form, differentiation, and meaning.

from non-conscious organisms that did not specify maximal φ_s as integrated wholes, yet were composed either of i) ontologically reducible collections of minimally conscious entities intrinsically existing at a lower scale of reality (e.g., an atomic or molecular 'ontological dust') (Tononi et al., 2022; Tononi & Boly, 2025), ii) formless stuff (Koch, 2024), or iii) a combination of both. In any case, unless one accepts an explanatory regress back to a primordial ontological dust, any conscious entity—whether a fully conscious organism or minimally conscious 'dust' (e.g., atoms or molecules) constituting non-conscious organisms—must ultimately have evolved or arisen from previously non-conscious, undifferentiated, formless stuff. We refer to Koch's proposal as the *formless stuff hypothesis*.

If this is the case, then what exists would not be completely exhausted by IIT's "Great Divide of Being": beyond the intrinsic/extrinsic division we would find a third category, what may be called "formless reality", that in contrast to extrinsic existence, is not grounded on the intrinsic existence of consciousness. Thus, according to this proposal, there could be something instead of nothing in a universe completely devoid of even minimally conscious intrinsic micro-entities.

Although the appeal of the formless stuff hypothesis is compelling, the following section will argue against it. More specifically, it will be argued i) against the (IIT-theoretic) plausibility of a third type of existence or reality beyond the intrinsic/extrinsic division, as well as ii) against the idea that in the absence of consciousness, reality is just an undifferentiated, unstructured ontological chaos. Finally, it is also argued that Koch's main motivation for positing this "formless stuff" rests on an epistemological-ontological conflation.

Refutation of the "formless stuff" hypothesis

First, the very suggestion that, in addition to intrinsic and extrinsic existence, there is a third type of reality, constituted by "formless stuff", is theoretically very unstable. According to the most straightforward reading of IIT, the 'Great Divide of Being' covers all that exists---dividing all reality into either intrinsic or extrinsic entities. Indeed, in his previous book, Koch himself writes that "causal power of two different kinds [i.e., intrinsic and extrinsic] is the only sort of stuff needed to explain everything in the universe. These powers constitute ultimate reality" (Koch, 2019, p. 166). Given that the difference between intrinsic and extrinsic causal power corresponds, in scientific-operational terms, to the difference between intrinsic and extrinsic existence, it means that Koch himself is claiming that the Great Divide of Being exhausts all reality, leaving *nothing* outside. Furthermore, this position is understandable given IIT's commitments: only conscious, intrinsic existence is true existence; while extrinsic existence is relative and a "lesser" kind of existence, but nonetheless different from nothing because it is grounded on the true existence of conscious experience (i.e., existence *for* an intrinsic entity). But if Koch's "formless stuff" is neither intrinsic nor extrinsic existence, then it is very difficult to see how it could be distinct from nothing without undermining the intended metaphysical foundation for a comprehensive view of nature that IIT attempts to offer with its "Great Divide".

Second, the very suggestion that without consciousness, there is only formless, undifferentiated stuff is deeply problematic for both the scientific worldview and common sense. It can be seen by recalling Koch's own example: "Think of the early Earth... the planet was lifeless... Despite the sun shining, Earth's surface was dark,

without the inner light of consciousness" (Koch 2024, p. 105). The inconsistency is that "the sun shining" is far from being a formless, unstructured stuff. More generally, one of the central aims of science is precisely to describe and explain the order and structure of the world. Consequently, the scientific view of the non-conscious world is not of undifferentiated chaos; this is precisely what Koch suggests when he writes that before the origin of consciousness on the early Earth, the sun still shone: there was some significant, intelligible order, as inferred by planetary science, paleoclimatology, and geophysics, among others. In the absence of strong reasons to the contrary, it seems more reasonable to assume that this order existed prior to the "first flickering of phenomenal light" (Koch 2024, p. 105), even if there was nobody who could observe and appreciate the meaning and value of such an order. Claiming otherwise seems to conflate the objective existence of order and structure in nature, with our structured experience and knowledge of it.

Now, it is important to note that we agree with Frank et al. (2024) that the mainstream scientific worldview is incomplete, suffering from the blind spot that ignores the role of consciousness in the creation of scientific knowledge. This omission leads to a fragmented understanding of reality, in which the knower is absent from the known, creating a significant vacuum in a worldview that aims to be comprehensive but overlooks the very experience through which knowledge is acquired. Nonetheless, this is different from Koch's radical view, which suggests that without consciousness, there is no coherent and structured existence at all —no trees, no forests, just 'formless stuff'. While Frank et al. (2024) argue for acknowledging consciousness as an essential component in understanding, creating, and integrating scientific knowledge, they do not go as far as claiming that the complexity, structure, and coherence of physical reality itself are entirely contingent upon consciousness. Also, while it seems to be almost tautologically true that consciousness is required for the world to appear *as we experience it* (i.e., our experiential or phenomenal world), this does not entail that consciousness is required for the world to exist in itself as an incredibly complex, structured and lawful web of interacting processes and components at multiple spatial and temporal scales (Capra, 1997; Capra & Luisi, 2014; Thompson, 2007).

Thus, although we find the appeal of Kant's insight that we may never be able to go completely beyond experience and grasp reality *as it is in itself* (i.e., as *noumena* beyond *phenomena*), this does not entail that reality beyond consciousness is completely deprived of order, regularity, and structure. This is because Kant's insight is fundamentally an epistemological claim (concerning the limits of our knowledge), while the latter is an ontological or metaphysical one, which does not necessarily follow from the former. Koch seems to conflate the epistemological and phenomenological necessity of consciousness regarding our knowledge and experience of the world with an ontological claim that denies the independent existence of a complexly organized reality beyond consciousness.

This *epistemological-ontological conflation* is most salient in Koch's claim that without a conscious observer there are neither trees nor a forest, "as these *concepts* depend on a conscious subject discerning a tree from other trees, treating them as different from the soil...Nature knows nothing of these distinctions... only of formless stuff" (2024, p. 105). To argue that trees and forests don't exist without consciousness because the *concepts* of trees and forests don't exist without consciousness, is to confuse epistemology with ontology, at least under a realist understanding of "ontology". On such a view, ontology as a philosophical discipline certainly depends on human knowledge, but its aim is precisely to designate and describe (and perhaps explain) the fundamental structure of

a reality that exceeds and is independent of our ontological concepts, much as the name 'Nubecita' refers to one of my cats, which exists beyond and independently of the name itself, even if the name is contingent on human language and the concept of my cat is mind-dependent.

One could alternatively interpret Koch as endorsing a nominalist view, according to which categories such as trees or forests do not exist as natural kinds independently of conscious beings. However, our concern does not hinge on realism about such categories, but on the apparent requirement that the evolutionary emergence of consciousness presupposes a sufficiently structured and ontologically robust pre-conscious reality. Otherwise, it becomes difficult to make sense of how consciousness could emerge from organisms whose own existence is not granted independent ontological status. For the Sun to shine for pre-sentient organisms, the world in which they evolved must already possess real structure and regularity independently of consciousness, rather than being undifferentiated stuff.

Thus, Koch incorrectly extends an *epistemological* claim about our conceptual understanding, categorization, and perception of trees and forests, about our "ontology" of the world as a form of human knowledge, into an *ontological* claim about the actual properties and existence of trees and forests. But it is far more consistent with both scientific knowledge and common sense to state, that without consciousness, a forest is still a complexly structured ecosystem with many interacting components and processes, even if there is no one to appreciate it.

Indeed, positing a complex, structured reality before the origin of consciousness on Earth offers a more coherent explanation for the evolutionary development of consciousness than the idea of a pre-existing formless ontological chaos. Acknowledging the reality of intricate interactions between non-conscious entities, both in the context of cosmological and evolutionary history, allows for a better understanding of how consciousness might have evolved naturally from this pre-existing complex and lawfully structured reality, rather than mysteriously arising from a state of undifferentiated, meaningless, and "formless" stuff. The latter account portrays consciousness as a sudden, inexplicable leap from something fundamentally different and almost unintelligible, contradicting the gradual change expected in evolutionary explanations.

Concluding remarks

If IIT retains the view that consciousness grounds all forms of existence (i.e., intrinsic and extrinsic), to account for the origin of consciousness, the theory ultimately needs to posit a primordial ontological dust (i.e., a "Big-Bang consciousness"), which is unfortunately a very problematic proposal (Cea et al., 2024a). Alternatively, IIT may relax its "Great Divide of Being" and claim that beyond the intrinsic-extrinsic division, there is something instead of nothing, an unstructured reality inhabited by "formless stuff" (Koch, 2024). However, it is argued i) that it is not (IIT-theoretically) sound to posit a third kind of existence beyond the intrinsic/extrinsic distinction, and ii) that, even in the absence of consciousness, reality is not an undifferentiated, unstructured ontological chaos. Finally, it is also maintained iii) that Koch's main motivation for introducing this "formless stuff" rested on an epistemological-ontological conflation, in which the necessity of our concepts and categories of objects and relations for knowing and experiencing a structured reality is illegitimately conflated with the (unsupported) necessity of those concepts and categories for reality itself to exist as structured.

So, what could IIT's way out of this conundrum be? As argued by Cea & Signorelli (2025), IIT should adopt a form of *causal-physical realism*, that requires three adjustments. First, to reject the principle of true existence (PTE) and its associated Great Divide of Being. Second, to revise IIT's principle of maximal existence in favor of the principle of maximal *conscious* existence: "what exists consciously is what holistically exists the most" (PMCE). "Holistically" means that systems instantiating $\varphi_s > 0$ truly exist as causally irreducible wholes even when they do not maximize φ_s , while systems instantiating maximal φ_s have, in addition, the property of being conscious. Finally, IIT should revise its principle of being in favor of the principle of *realistic* being ("PRB"), according to which "to truly exist is to have causal power". In IIT's original principle of being ("PB"), physical existence is defined operationally as the capacity to take and make a difference, yet true existence is reserved for conscious intrinsic entities, which forces the principle to be interpreted in a merely instrumental sense. The PRB removes this restriction by holding that causal power itself—which in IIT is quantified in terms of *intrinsic information ii*—suffices for genuine existence (Albantakis et al., 2023). On this view, even systems that do not instantiate $\varphi_s > 0$, and thus do not exist as irreducible wholes, let alone as conscious entities, may still exist insofar as they instantiate $ii > 0$ and thus possess intrinsic cause-effect power (Cea et al., 2025; Cea & Signorelli, 2025).

Taken together, these commitments would enable IIT theorists to continue using their current neuroscientific methodology and computational framework to identify the physical substrate of consciousness (i.e., the complex) and to unfold its cause-effect structure, without thereby implying that non-conscious entities lack mind-independent, ordered, genuine existence.

In this way, realism about a world that precedes and extends beyond any consciousness is vindicated, and the received view that consciousness phylogenetically originated from non-conscious living ancestors can be straightforwardly accommodated, while also resolving other difficulties that stem from equating true existence with subjective experience (Cea et al., 2024b; Signorelli et al., 2025). More generally, alternative metaphysical frameworks compatible with IIT's formalism may also be considered, such as forms of emergentism (Cea, 2021; Negro, 2022b) that depict the world as a complex, multilevel network of mutually affecting powerful entities, of which only a subset might be conscious.

References

- Albantakis, L., Barbosa, L., Findlay, G., Grasso, M., Haun, A. M., Marshall, W., Mayner, W. G. P., Zaeemzadeh, A., Boly, M., Juel, B. E., Sasai, S., Fujii, K., David, I., Hendren, J., Lang, J. P., & Tononi, G. (2023). Integrated information theory (IIT) 4.0: Formulating the properties of phenomenal existence in physical terms. *PLOS Computational Biology*, 19(10), e1011465. <https://doi.org/10.1371/journal.pcbi.1011465>
- Albantakis, L., Hintze, A., Koch, C., Adami, C., & Tononi, G. (2014). Evolution of integrated causal structures in animats exposed to environments of increasing complexity. *PLoS Computational Biology*, 10(12), e1003966.
- Birch, J. (2024). *The Edge of Sentience: Risk and Precaution in Humans, Other Animals, and AI*.
- Capra, F. (1997). *The web of life: A new scientific understanding of living systems*. Anchor.
- Capra, F., & Luisi, P. L. (2014). *The systems view of life: A unifying vision*. Cambridge University Press.
- Cea, I. (2021). Integrated Information Theory of Consciousness is a Functionalist Emergentism. *Synthese*, 199(1-2), 2199-2224. <https://doi.org/10.1007/s11229-020-02878-8>

- Cea, I. (2023a). On motivating irruptions: the need for a multilevel approach at the interface between life and mind. *Adaptive Behavior*.
<https://doi.org/10.1177/10597123231184651>
- Cea, I. (2023b). *The Somatic Roots of Affect: Toward a Body-Centered Education*. In Fossa, P., & Cortés-Rivera, C. (Eds.), *Affectivity and Learning: Bridging the Gap Between Neurosciences, Cultural and Cognitive Psychology* (pp. 555-583). Springer.
- Cea, I., & Martínez-Pernía, D. (2023). Continuous Organismic Sentience as the integration of Core Affect and Vitality. *Journal of Consciousness Studies*, 30(3-4), 7-33.
<https://doi.org/10.53765/20512201.30.3.007>
- Cea, I., Negro, N., Mediano, P. A. M., & Signorelli, C. M. (2025). *Phi-monads are conceptually problematic and mathematically undefined in Integrated Information Theory 4.0*. PsyArXiv.
https://doi.org/10.31234/osf.io/qm2bd_v1
- Cea, I., Negro, N., & Signorelli, C. M. (2023). The Fundamental Tension in Integrated Information Theory 4.0's Realist Idealism. *Entropy*, 25(10), 1453. <https://doi.org/10.3390/E25101453>
- Cea, I., Negro, N., & Signorelli, C. M. (2024a). *Big Bang consciousness: IIT 4.0 and the origin of subjective experience*. PsyArXiv.
<https://doi.org/10.31234/osf.io/q9zmr>
- Cea, I., Negro, N., & Signorelli, C. M. (2024b). Only Consciousness Truly Exists? Two Problems for IIT 4.0's ontology. *Frontiers in Psychology*, 15. <https://doi.org/10.3389/fpsyg.2024.1485433>
- Cea, I., & Signorelli, C. M. (2025). How to be an integrated information theorist without losing your body. *Frontiers in Computational Neuroscience*, 18(18), 1510066.
- Chalmers, D. (2021). *The routledge Handbook of idealism and immaterialism*. Routledge. 591-613.
- Chis-Ciure, R. (2022). The transcendental deduction of Integrated Information Theory: connecting the axioms, postulates, and identity through categories. *Synthese*, 200(3), 1-27.
- Cleeremans, A., Mudrik, L., & Seth, A. K. (2025). Consciousness science: where are we, where are we going, and what if we get there? *Frontiers in Science*, 3(3).
<https://doi.org/10.3389/fsci.2025.1546279>
- Comolatti, R., Grasso, M., & Tononi, G. (2025). Why does time feel the way it does? Toward a principled account of temporal experience. *IScience*, 28(10), 113434.
<https://doi.org/10.1016/j.isci.2025.113434>
- Consortium, C. (2025). Adversarial testing of global neuronal workspace and integrated information theories of consciousness. *Nature*, 1-10.
- Damasio, A., & Damasio, H. (2024). Homeostatic feelings and the emergence of consciousness. *Journal of Cognitive Neuroscience*, 36(8), 1653-1659.
- Dehaene, S. (2014). *Consciousness and the brain: Deciphering how the brain codes our thoughts*. Penguin.
- Dehaene, S., & Naccache, L. (2001). Towards a cognitive neuroscience of consciousness: Basic evidence and a workspace framework. *Cognition*, 79(1-2), 1-37.
[https://doi.org/10.1016/S0010-0277\(00\)00123-2](https://doi.org/10.1016/S0010-0277(00)00123-2)
- Denton, D. (2005). *The primordial emotions: The dawning of consciousness*. Oxford University Press.
- Denton, D., McKinley, M. J., Farrell, M., & Egan, G. F. (2009). The role of primordial emotions in the evolutionary origin of consciousness. *Consciousness and Cognition*, 18(2), 500-514.
- Ellia, F., & Chis-Ciure, R. (2022). Consciousness and complexity: Neurobiological naturalism and integrated information theory. *Consciousness and Cognition*(100), 103281.
- Ellia, F., Hendren, J., Grasso, M., Kozma, C., Mindt, G., Lang, P., Haun, M., Alabantakis, L., Boly, M., & Tononi, G. (2021). Consciousness and the fallacy of misplaced objectivity. *Neuroscience of Consciousness*, 2021(2), 1-12.
<https://doi.org/10.1093/NC/NIAB032>
- Feinberg, T. E., & Mallatt, J. (2013). The evolutionary and genetic origins of consciousness in the Cambrian Period over 500 million years ago. *Frontiers in Psychology*, 4(4), 667.
- Feinberg, T. E., & Mallatt, J. (2016). *The ancient origins of consciousness: How the brain created experience*. MIT Press.
- Frank, A., Gleiser, M., & Thompson, E. (2024). *The blind spot: why science cannot ignore human experience*. MIT Press.
- Fuchs, T. (2018). *Ecology of the Brain: The Phenomenology and Biology of the Embodied Mind*. Oxford University Press.

- Ginsburg, S., & Jablonka, E. (2019). *The evolution of the sensitive soul: learning and the origins of consciousness*. MIT Press.
- Godfrey-Smith, P. (2016). *Other minds: The octopus, the sea, and the deep origins of consciousness*. Farrar, Straus and Giroux.
- Godfrey-Smith, P. (2020). *Metazoa: Animal life and the birth of the mind*. Farrar, Straus and Giroux.
- Godfrey-Smith, P. (2024). *other*.
- Grasso, M., Juel, B. E., Hendren, J., & Tononi, G. (2024). *IIT Wiki*.
- Graziano, M. S. A. (2017). The attention schema theory: A foundation for engineering artificial consciousness. *Frontiers in Robotics and AI*, 4(4), 60.
- Graziano, M. S. A. (2019). *Rethinking consciousness: a scientific theory of subjective experience*. WW Norton & Company.
- Humphrey, N. (2022). *Sentience: The invention of consciousness*. Oxford University Press.
- Kant, I. (1999). *Critique of pure reason*. Cambridge university press.
- Kister, M. (2007). Tohu wa-Bohu, Primordial Elements and Creatio ex Nihilo. *Jewish Studies Quarterly*, 14(3), 229-256.
- Koch, C. (2019). *The Feeling of Life Itself: Why Consciousness is Widespread But Can't be Computed*. Mit Press.
- Koch, C. (2024). *Then I Am Myself the World: What Consciousness Is and How to Expand It*. Basic Books.
- Mashour, G. A., Roelfsema, P., Changeux, J. P., & Dehaene, S. (2020). Conscious Processing and the Global Neuronal Workspace Hypothesis. *Neuron*, 105(5), 776-798.
<https://doi.org/10.1016/J.NEURON.2020.01.026>
- Miller, A. (2021). *Realism*. In Zalta, E. N. (Ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2021).
- Negro, N. (2020). Phenomenology-first versus third-person approaches in the science of consciousness: the case of the integrated information theory and the unfolding argument. *Phenomenology and the Cognitive Sciences*, 19(5), 979-996.
- Negro, N. (2022a). Can the Integrated Information Theory Explain Consciousness from Consciousness Itself? *Review of Philosophy and Psychology*.
<https://doi.org/10.1007/s13164-022-00653-x>
- Negro, N. (2022b). Emergentist Integrated Information Theory. *Erkenntnis*, 1-23.
- Peebles, P. J. E. (2020). *Principles of physical cosmology*. Princeton university press.
- Searle, J. R. (1995). *The construction of social reality*. Simon and Schuster.
- Seth, A. (2021). *Being you: A new science of consciousness*. Penguin.
- Seth, A. (2024). *Interoception: A Comprehensive Guide*. Springer. 105-124.
- Seth, A. (2025). Conscious artificial intelligence and biological naturalism. *Behavioral & Brain Sciences*, 1-42.
<https://doi.org/10.1017/S0140525X25000032>
- Seth, A., & Bayne, T. (2022). Theories of consciousness. *Nature Reviews Neuroscience*, 23, 439-452.
- Signorelli, C. M., Cea, I., & Prentner, R. (2025). We need to explain subjective experience, but its explanation may not be mechanistic. *Neuroscience of Consciousness*, 2026(1), 2026.
<https://doi.org/10.1093/nc/niaf034>
- Signorelli, C. M., Szczotka, J., & Prentner, R. (2021). Explanatory profiles of models of consciousness-towards a systematic classification. *Neuroscience of Consciousness*, 2021(2).
- Storm, J. F., Klink, P. C., Aru, J., Senn, W., Goebel, R., Pigorini, A., Avanzini, P., Vanduffel, W., Roelfsema, P. R., & Massimini, M. (2024). An integrative, multiscale view on neural theories of consciousness. *Neuron*, 112(10), 1531-1552.
- Thompson, E. (2007). *Mind in Life: Biology, Phenomenology, and the Sciences of Mind*. Harvard University Press.
- Thompson, E. (2015). *Waking, dreaming, being: Self and consciousness in neuroscience, meditation, and philosophy*. Columbia University Press.
- Tononi, G. (2017). *Integrated information theory of consciousness: Some ontological considerations*. In Schneider, S., & Velmans, M. (Eds.), *The Blackwell companion to consciousness* (2nd edition) (pp. 621-633). Wiley Online Library.
<https://doi.org/10.1002/9781119132363.ch44>

- Tononi, G., Al bantakis, L., Boly, M., Cirelli, C., & Koch, C. (2022). *ArXiv*.
<https://arxiv.org/abs/2206.02069>
- Tononi, G., & Boly, M. (2025). *Integrated Information Theory: A Consciousness-First Approach to What Exists*. *ArXiv*.
<https://arxiv.org/abs/2510.25998>
- Tononi, G., & Koch, C. (2015). Consciousness: Here, there and everywhere? *Philosophical Transactions of the Royal Society B: Biological Sciences*. <https://doi.org/10.1098/rstb.2014.0167>
- Veit, W. (2023). Complexity and the evolution of consciousness. *Biological Theory*, 18(3), 175-190.