Big Bang consciousness: IIT 4.0 and the origin of subjective experience

Ignacio Cea[1,2*], Niccolo Negro[3], Camilo Miguel Signorelli[4,5,6]

[1]Center for Research, Innovation and Creation, and Faculty of Religious Sciences and Philosophy, Temuco Catholic University, Temuco, Chile

[2]Philosophy Department, Faculty of Philosophy and Humanities, Universidad Alberto Hurtado, Santiago, Chile

[3]School of Psychological Sciences, Tel Aviv University, Tel Aviv, Israel

[4]Department of Computer Science, University of Oxford, Oxford, United Kingdom

[5]Center for Philosophy of Artificial Intelligence, University of Copenhagen, Copenhagen, Denmark

[6]Laboratory of Neurophysiology and Movement Biomechanics, Université Libre de Bruxelles, Brussels, Belgium

*** Correspondence:**
Ignacio Cea
igneocj@gmail.com

Abstract

This paper addresses a problem that arises from the ontological commitments of Integrated Information Theory (IIT) 4.0, particularly its stance that only conscious entities truly exist. This position leads to the "origin of consciousness problem": if non-conscious entities do not truly exist, how could consciousness have evolved from non-conscious ancestors? We explore several responses IIT might offer, such as the co-origin of life and consciousness, or the idea that non-conscious ancestors might have been constituted by "ontological dust"—minimally conscious, intrinsic micro-entities collectively aggregated to form bigger objects lacking unified consciousness. Our analysis shows that IIT's ontological framework, along with scientific knowledge regarding biological evolution, prebiotic chemical structures, and

physical cosmology, ultimately forces the theory into positing a form of "Big Bang consciousness", that is, a *primordial ontological dust* constituted by minimally conscious elementary particles created soon after the Big Bang. Although IIT may accept this striking implication, we think that it introduces tensions with both the received scientific view of the evolutionary origin of consciousness and the cosmological understanding of early universe components. We also present but ultimately reject an alternative option based on what we call the “formless stuff hypothesis”, which might avoid the implication that consciousness originates from nothing as well as the necessity of a "Big Bang consciousness”. We conclude by suggesting that IIT's metaphysical commitments, especially the equation true existence=phenomenal existence, require re-examination to reconcile its framework with standard scientific knowledge, and in particular, with the received view about the phylogenetic origin of consciousness.

Keywords

Integrated Information Theory; Evolution of consciousness; Phenomenal existence; Ontological dust; Cosmology of consciousness; Great Divide of Being

1. Introduction: IIT 4.0 essentials

Integrated Information Theory (IIT) is currently regarded as one of the foremost scientific approaches to understanding consciousness (Consortium et al., 2023; A. K. 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[1] 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, Tononi 2022). 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 it claims are essential to consciousness: intrinsicality, 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 instantiate consciousness. The theory mathematically formalizes these postulates, identifying the "complex", also called the physical substrate of consciousness, as the system that achieves the *maximal* value of *system integrated*

[1] Throughout this article, we use the term "consciousness" according to the phenomenal sense, i.e., as subjective experience (Block, 2007; Chalmers, 1996; Nagel, 1974). Also, sometimes we use "sentience" interchangeably with "consciousness", although strictly speaking, sentience refers to a subset of phenomenally conscious states: affective states imbued with hedonic valence, often associated with feelings, and emotions, but also with felt physiological drives such as hunger or thirst (Birch, 2024; Cea & Martínez-Pernía, 2023; Denton, 2005).

information (φ_s*), while "overlapping substrates with lower φ_s are *excluded from existence*" (Albantakis et al., 2023, p. 12, italics added)[2].

In IIT, this "exclusion from existence" is literal, not metaphorical. IIT posits that candidate substrates that do not specify the maximal value of system integrated information (φ_s*) do not truly exist, because they do not exist "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, if we were to enumerate the entities that truly exist in objective reality — meaning those that exist in themselves, independently of external observers— we would need to include only those physical systems that intrinsically exist as subjective experiences (Cea et al., 2023)[3].

This is the foundation for IIT's "Great Divide of Being" (Koch, 2024; Tononi, 2017; Tononi et al., 2022). The latter 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. Yet, can this extrinsic existence still be considered a proper form of existence, "lesser" than intrinsic existence but distinct from

[2] In the IIT 4.0 formalism, there are other important measures like the *integrated information for distinctions φ_d, integrated information for relations φ_r, structure integrated information Φ,* and the *Φ-structure.* However, given the scope of this article, we focus on the *system integrated information φ_s* , and in particular, its maximal value among overlapping systems (φ_s*). We refer the reader to Albantakis et al. (2023) for details about the other measures and the whole IIT 4.0 formalism.

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

non-existence? Koch suggests this when he states that overlapping substrates "of lower integrated information exist, but only extrinsically, for others, not for themselves" (2024, p. 104). Nonetheless, given IIT's attribution of *true* existence *only* to intrinsic existence (Albantakis et al., 2022; Cea et al., 2023; Hendren et al., 2024a; Tononi et al., 2022), and that extrinsic existence is *not* intrinsic existence, logic indicates that extrinsic existence is *not true* existence. Moreover, if "not true" = "false", then extrinsic existence would be *false* existence, or non-existence (i.e. nothing).

However, this analysis may obscure the more nuanced meaning of "extrinsic existence" that is intended by IIT. According to the theory, "extrinsic existence is grounded in intrinsic existence… extrinsic interactions can be considered properties of intrinsic existence" (Hendren et al., 2024b, fn. 3). 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), as well as to an *eliminative idealism* (i.e., denies *mind-independent* existence) toward all physical entities that do not maximize φ_S and hence are non-conscious, including familiar large-scale objects like bodies and stars. Importantly, this is conceptually compatible with Koch's claim that extrinsic

entities do exist, but only if their existence is considered mind-dependent, i.e. extrinsic entities do not exist independently of some other conscious, intrinsic entity that experiences them. Thus, the existence of extrinsic entities seems to be analogous to that of social kinds like money or governments, mind-dependent but nonetheless different from nothing (Searle 1995).

To clarify, 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 my 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).

Thus, IIT is realist, in the standard sense of the term, only about intrinsic entities, since extrinsic entities do not exist independently of another intrinsic entity's perspective. Therefore, IIT's external-world realism should be understood as claiming that an external world, beyond my own consciousness, truly exists, but is fundamentally constituted by other conscious, intrinsic entities, each one enjoying their own subjective experience. This can be understood as a form of *realist idealism* (Cea et al., 2023; Chalmers, 2021).

In the remaining sections, we present an important problem that follows from these ontological assumptions: the "origin of consciousness problem". This is the question about how consciousness could have evolved from non-conscious ancestors if, according to IIT, without consciousness there is neither intrinsic nor extrinsic existence. We show that to address this issue, IIT is ultimately entitled to posit a *primordial ontological dust* or "Big

Bang consciousness": a set of minimally conscious, elementary particles created soon after the Big Bang.

2. 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; Ginsburg & Jablonka, 2019; Godfrey-Smith, 2016; Veit, 2023). In other words, non-conscious forms of life eventually evolved and developed primordial forms of consciousness, probably due to the emergence of the right sort of nervous systems both in vertebrate and invertebrate animals (Birch, 2024; Damasio & Damasio, 2024; Denton, 2005; Godfrey-Smith, 2024; Humphrey, 2022).

Thus, according to what we might call 'the received view,' consciousness phylogenetically evolved from certain non-conscious processes in ancestors that lacked sentience; that is, the first sentient organisms evolved from previously non-conscious living ancestors. 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). 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).

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 received view is right, and there was a time on Earth when there was no consciousness at all, then our non-conscious ancestors did not exist extrinsically either, because there was no intrinsic entity for them to exist for. 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 we call 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.

In the following section, we will explore IIT's potential ways of dealing with this problem. We will argue that all viable possibilities lead IIT to postulate the existence of what we will

call *primordial ontological dust*: a collection of intrinsically existing, minimally conscious, elementary particles without internal structure that emerged in the earliest moments of our universe, i.e., a kind of "Big Bang (minimal) consciousness".

3. All roads lead to a primordial ontological dust

When faced with the origin of consciousness problem, IIT seems to be forced into the following trilemma[4], according to which only one of these three options can be right:

i) CONSCIOUSNESS FROM NOTHING: If IIT accepts the received view, then a possibility is that the theory could simply accept that consciousness evolved from non-conscious organisms (e.g. Precambrian organisms without nervous systems), that didn't exist because they were neither intrinsic nor extrinsic entities (i.e. they didn't even exist from the perspective of other intrinsic entities, because there was none). This option accepts the counterintuitive implication of the origin of consciousness problem for IIT: conscious experience originated from nothing at some point in what biology considers the evolution of life on Earth;

ii) CONSCIOUSNESS-LIFE CO-ORIGIN: Alternatively, IIT may reject the received view, and assert that consciousness didn't originate at some point in the evolution of life, but was present –albeit in very basic forms– from the beginning of life itself and gradually evolved because of the increased complexity of organisms. This alternative dissolves the origin of consciousness problem as we stated it, but puts IIT in tension with a widespread scientific

[4] We are disregarding here a fourth possibility: the presence of an all-knowing universal consciousness, from the perspective of which non-conscious ancestors and their environments may have existed as extrinsic entities.

view about the phylogenetic emergence of consciousness, what we called the received view. Moreover, as we will analyze below (section 3.2.), it creates a new problem, what we may call "the origin of life & consciousness problem";

iii) CONSCIOUSNESS FROM ONTOLOGICAL DUST: Finally, if IIT prefers to align with the received view, while also holding that that there was *something* before the origin of biological forms of consciousness, then IIT can maintain that consciousness originated from non-conscious organisms, stipulating that those ancient organisms existed as reducible aggregates of minimally-conscious, intrinsically existing parts. The latter would be micro-units of intrinsic existence that specify small –albeit non-zero– values of φ_s*, and which IIT refers to as "ontological dust" (Tononi et al., 2022). These "particles" of ontological dust may collectively constitute bigger objects such as nonconscious multicellular organisms. Thus, according to this third option, the non-conscious ancestors of the first sentient creatures lacked a unitary conscious experience, and hence were not intrinsic entities, but nonetheless were constituted by a multiplicity of minimally conscious parts (e.g., intrinsically existing carbon molecules).

Now, we address options i) and ii). Option iii) will be analyzed afterwards in section 3.3.

3.1. CONSCIOUSNESS FROM NOTHING (option i)

We claim that option i) can be disregarded given the almost contradictory nature of claiming that something that didn't exist gave rise to something that truly exists in itself and for itself. This would amount to asserting that consciousness magically came to existence from, literally speaking, nothing. We do not take IIT to be committed to such a contradictory claim,

since IIT proponents are aware that it is impossible to make something out of nothing (Albantakis et al. 2023). Moreover, this would entail the rejection of the evolutionary story of life on Earth, because embracing this option means that organisms at a pre-consciousness stage of evolution didn't exist. In other words, this alternative entails the awkward conclusion that the first creatures that ever existed were the ones satisfying the conditions for consciousness (e.g., those with nervous systems sufficiently integrated). Hence, disregarding the presence of an all-knowing universal consciousness at that time, from the perspective of which non-conscious ancestors and their environments may have existed as extrinsic entities, option i) entails that the Earth and its inhabitants really came into existence only when suitably equipped conscious organisms appeared. Thus, we take option i) to be too implausible given its utterly extravagant consequences.

3.2. CONSCIOUSNESS-LIFE CO-ORIGIN (option ii)

According to option ii), consciousness was present even in LUCA (Last Universal Common Ancestor), the single-celled organisms that are believed to be the common ancestors of all species on Earth (Weiss et al., 2016). Thus, this option entails the rejection of what we called the received view, i.e., that consciousness first evolved from non-conscious processes in living ancestors lacking subjective experience.

In relation to the prospects of cellular consciousness from the perspective of IIT, a relevant study is Marshall et al. (2017), in which IIT's formal analysis is applied to the cell-cycle model of a unicellular eukaryote, the fission yeast (*Schizosaccharomyces pombe*), revealing

that it specifies a global maximum of integrated information Φ=0.43 (Marshall et al 2017)[5]. Thus, it is not unreasonable to suppose that applying IIT to a model of LUCA may reveal some positive amount of integrated information as well.

Importantly, if this route is embraced by IIT, then the theory would be committed to a form of *biopsychism*. In this view, all life has a mind (Haeckel, 1892). Early advocates of this view are the French psychologist Alfred Binet (1889) and American zoologist Herbert Jennings (1931), according to whom sentience is present in all living organisms. More recently, this view is embraced by Margulis and Sagan (2000), who claim that "every organic being, every autopoietic cell is conscious" (p. 150), or Reber et al. (2023), according to which "all life is sentient. Life and consciousness are coterminous" (p. 1).

This radical position, however, is not even embraced by some of the fiercest defenders of the life-mind continuity thesis (Godfrey-Smith, 1998, 2024; Thompson, 2007, 2022), nor by those who see consciousness as fundamentally linked to life-regulation and survival (Damasio & Damasio, 2023, 2024; Denton et al., 2009; A. Seth, 2021). One of the main reasons to reject that view is the common opinion that nervous systems are necessary for (biological) consciousness[6], something which is even apparently held by IIT itself (Tononi et al. 2022, Albantakis et al. 2014, Koch 2024). Hence, embracing this option would make IIT able to avoid the implication of option i) that consciousness evolved from non-conscious processes in non-existent, non-conscious predecessors, but with the cost of having to reject the received view: the necessity of nervous systems for biological consciousness and the evolution of conscious life from non-conscious life.

[5] However, the value is computed using the IIT 3.0 formalism (Oizumi et al., 2014), which does not distinguish between system integrated information (φ_S), the metric we are interested in here, and structure integrated information (Φ), as IIT 4.0 currently does (Albantakis et al. 2023).
[6] However, we remain neutral concerning this widespread view about the necessity of a suitably integrated or complex nervous system for consciousness.

Furthermore, and even more importantly, option ii) brings about a new problem for IIT, what we may call "the origin of life & consciousness problem": How could sentient life have first evolved from non-sentient chemistry, if according to IIT nothing exists without consciousness? To face this problem, a further trilemma opens up for IIT, depending on how one interprets the origin of LUCA as a conscious entity. Here are three possible options:

ii-A) CONSCIOUS-LUCA FROM NOTHING: Conscious LUCA originated from *non-conscious*, prebiological chemical structures (e.g. autocatalytic sets, Kauffman (1993)) that neither existed for themselves, nor for other intrinsic entities, i.e., they were nothing;

ii-B) CONSCIOUS-LUCA FROM CONSCIOUS CHEMISTRY: Conscious LUCA originated from *conscious*, prebiological chemical structures that truly existed in themselves (e.g. minimally-conscious prebiotic autocatalytic networks);

ii-C) CONSCIOUS-LUCA FROM ONTOLOGICAL DUST: Conscious LUCA originated from *non-conscious*, prebiological chemical structures that nonetheless existed as reducible aggregates of an "ontological dust" of minimally-conscious, intrinsically existing parts (as in the alternative iii) above).

Let us analyze options ii-A) and ii-B).

3.2.1. CONSCIOUS-LUCA FROM NOTHING (option ii-A)

Option ii-A) may be disregarded for the same reasons as i) above (section 3.1). In other words, it problematically entails that something non-existent could give rise to something that truly exists. But it seems impossible that nonexistent stuff could produce or give rise to something. Also, it implies that the Earth and both its biological and nonbiological inhabitants only came into existence when the first sentient/living organisms originated from nothing. So, let's analyze ii-B) before turning to options iii) and ii-C), both of which depend

on the plausibility of the notion of an ontological dust and hence will be analyzed jointly (section 3.3 below).

3.2.2. CONSCIOUS-LUCA FROM CONSCIOUS CHEMISTRY (option ii-B)

Option ii-B) suggests that sentient LUCA evolved from *something rather than nothing.* This "something" should comply with the requirements for true existence in the IIT sense, and was therefore conscious. However, being LUCA the first form of life, these pre-existing conscious things could not be *living* entities. They could have been, for example, prebiological physico-chemical structures such as autocatalytic networks (Kauffman 1993). That is, option (ii-B) states that consciousness was present on Earth before the origin of life. Now, although to the best of our knowledge, the IIT formalism has not been applied to prebiological systems such as autocatalytic networks, the latter may specify some amount of system-φ given their self-sustaining and recursive character, and increasing connectivity and complexity in simulated evolution (Hordijk et al., 2010; Hordijk & Steel, 2017; Jain & Krishna, 1998).

However, the origin of consciousness problem still applies: the components of a highly integrated autocatalytic network must have existed individually as conscious, intrinsic entities before composing the network. Otherwise, they would not have truly existed (i.e., they would be nothing), and therefore could not have later become part of a larger conscious system. In the case of Kauffman's Binary Polymer Model of autocatalytic networks (1993), the components are polymers (e.g., RNA molecules) constituted as strings of two types of monomers (e.g., nucleotides). Now, even if some rationale might be found to attribute a certain measure of φ_s* to these systems, the logic we have used can be recursively applied

down to the fundamental components of matter, i.e., the elementary particles. In other words, for potentially φ_s-specifying autocatalytic networks to emerge, their constituent molecules must have specified φ_s max individually before being part of those networks, but for these complex molecules to have emerged at some point in the molecular evolution of Earth, their simpler constituents (e.g., carbon atoms) must have also previously been intrinsic entities. This applies to their own components and so on, back to the appearance of the fundamental components of matter moments after the Big Bang. Hence, option ii-B) (CONSCIOUS-LUCA FROM CONSCIOUS CHEMISTRY) ultimately entails the existence of what we may call *primordial ontological dust*[7]: a set of intrinsically existing, minimally conscious elementary (partless) particles created at the earliest stages of our universe, i.e., a kind of "Big Bang consciousness".

3.3. CONSCIOUSNESS FROM ONTOLOGICAL DUST & CONSCIOUS-LUCA FROM ONTOLOGICAL DUST (options iii & ii-C)

As we will see now, the implication of a *primordial* ontological dust also follows from the remaining options iii) and ii-C): IIT may argue that consciousness evolved from either non-conscious organisms, or non-conscious prebiotic physico-chemical structures, both of which could have existed independently of external observers, albeit as reducible aggregates of "ontological dust". From the perspective of IIT, the notion of ontological dust is thus crucial to ensure that consciousness did not evolve from nothing.

[7] The adjective "primordial" is to specify that the constituents of this ontological dust are elementary entities residing at the ground-level of reality, and which were the first entities that ever existed. This is different from the broader notion of "ontological dust" (Tononi et al. 2022) which is scale-agnostic (i.e. can be constituted by intrinsically existing neurons, molecules, atoms, subatomic particles, etc.), and could exist at any point in time.

Let us explain these alternatives in more detail. Non-conscious organisms or prebiotic structures would be constituted by intrinsically existing parts such as minimally conscious molecules, atoms, or elementary particles. In fact, a similar hypothesis is suggested by Tononi and colleagues (2022), when they state that an unconscious body, which does not exist for itself, might be "just an aggregate of much smaller [intrinsic] entities" (p. 8). It is also suggested by Larissa Albantakis, who states that every physical entity may be considered either part of a conscious intrinsic entity, or a conscious intrinsic entity in itself (Albantakis, 2023).

That is, even if primordial organisms or prebiotic chemical structures were not conscious as such, because they didn't specify maximal φ_s, it is theoretically possible that they could have been constituted by minimally conscious parts, such as different sorts of molecules, atoms, or elementary particles/fields that may have individually specified maximal φ_s at a lower scale of nature, and hence existed intrinsically. The exact scale or "grain" at which system-φ is maximized, is in principle measurable (Hoel et al., 2016; Marshall et al., 2018), and according to IIT, it determines the grain at which an entity truly exists (Albantakis et al. 2023, Tononi et al. 2022). Thus, even if an autocatalytic network did not specify maximal φ_s as a whole, it could nonetheless be "condensed" (the term is from Albantakis et al 2023, p. 19) into an exhaustive and non-overlapping set of maximal φ_s molecules or atoms.

Let us further clarify the difference between a conscious intrinsic entity that specifies maximal φ_s as a whole, and a non-intrinsic entity that nonetheless may be said to exist objectively (i.e., not just from the perspective of an another intrinsic entity). While a φ_s-maximal intrinsic entity qualifies as *one* intrinsic entity with *one* unitary consciousness, a

non-intrinsic entity that nonetheless may exist objectively is a reductive aggregate of (minimally conscious) intrinsic entities that truly exist at a lower scale (i.e., a collection of "ontological dust"). In other words, an objectively existing non-intrinsic entity has smaller intrinsic entities as its parts but does not exist intrinsically as a whole beyond these parts. It is nothing but the "sum" of its constituents: it is a collection of *many* intrinsic micro-entities, each one with its own subjective experience, without a global unitary consciousness. In short, an intrinsic entity is *one* (i.e., one integrated consciousness)[8], while a non-intrinsic entity that nonetheless exists objectively is *many* (i.e. a plurality of consciousnesses).

However, even if a non-conscious, non-φ_s-maximal, autocatalytic network could have existed objectively as a reducible collection of *many* intrinsic micro-entities such as minimally conscious RNA molecules (or entities at a lower scale), for these RNA molecules to come to existence at some point in the early physico-chemical history of the Earth, their components (i.e. nucleotides and phosphodiester bonds) must have existed before. Either as intrinsic entities in themselves, or as aggregates of intrinsic entities residing at a lower scale (e.g. minimally conscious atoms of carbon, hydrogen, oxygen, etc.). But astrophysics and cosmology tell us that even carbon or hydrogen atoms on Earth were formed from simpler, pre-existing components, such as protons and neutrons inside the nuclei of stars like our Sun, and electrons that originated in the early universe. IIT can account for this by maintaining that these subatomic particles (plus force-carrying bosons like photons and gluons) were either, individually taken, global maxima of φ_s, or were aggregates of intrinsic entities residing at a still lower scale, e.g. aggregates of minimally conscious quarks (in the case of

[8] This is also supported by the view that in conscious intrinsic entities, the parts are "fused-up" into the whole, and cease to exist as individual entities on their own (Negro, 2022b).

protons and neutrons). But in either case, IIT ultimately needs to posit the existence of what we called a *primordial ontological dust*: minimally conscious elementary units of intrinsic existence residing at the most fundamental level of nature and created at the very beginnings of our universe.

3.4. CONSCIOUSNESS FROM PRIMORDIAL ONTOLOGICAL DUST

To summarize, we have achieved a converging result by analyzing both the previous, initially promising options ii) CONSCIOUSNESS-LIFE CO-ORIGIN and ii-B) CONSCIOUS-LUCA FROM CONSCIOUS CHEMISTRY, as well as options iii) CONSCIOUSNESS FROM ONTOLOGICAL DUST and ii-C) CONSCIOUS-LUCA FROM ONTOLOGICAL DUST. All options converge into the notion of a *primordial ontological dust,* namely, a set of fundamental-level, elementary (partless, indivisible) constituents of nature that were the first intrinsic entities that appeared after the Big Bang, and specified minimal, but non-zero, amounts of maximal φ_s and consciousness (i.e. “Big Bang consciousness”). We have summarized these results in Figure 1.

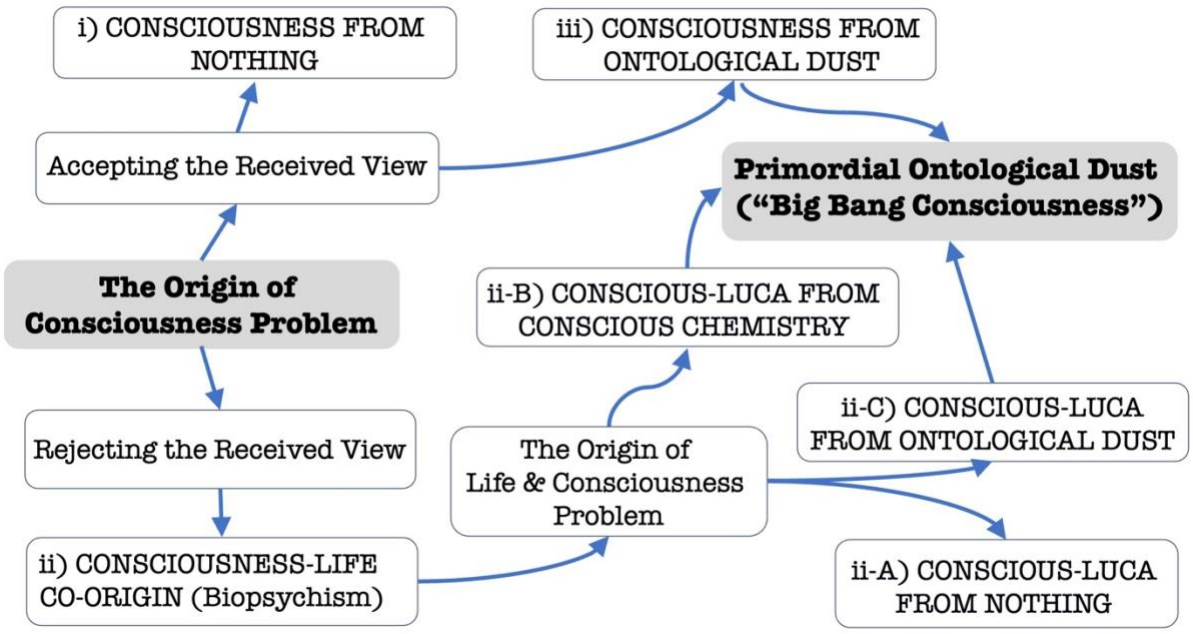

Figure 1. All roads lead to (either nothing or) a *primordial ontological dust.* Given IIT's ontological commitments–most centrally, its "Great Divide of Being–to account for the origin of consciousness, IIT ultimately needs to posit a *primordial ontological dust*: a set of minimally conscious elementary units of intrinsic existence residing at the most fundamental level of nature and created at the very beginnings of our universe, i.e., a sort of "Big Bang consciousness".

Thus, IIT's reply to the origin of consciousness problem might be that the received view about the evolution of biological consciousness from non-conscious life is correct, but it cannot be the whole story: the origin of consciousness needs to be supplemented with the notion of a (primordial) ontological dust:

"As the universe evolves, what first exists are merely small aggregates of ontological "dust"—intrinsic entities characterized by negligibly small Φ-structures. Organisms

eventually evolve with neural substrates, portions of which unfold into Φ-structures of high Φ." (Tononi et al. 2022, p. 21)

In other words, IIT's answer to the origin of consciousness problem ultimately rests on the assertion that whatever the final scientific account may reveal about the onset of consciousness in biological organisms on Earth, previous non-conscious entities (whether they are non-conscious primordial organisms, prebiotic chemistry, atoms, etc.) should be understood as being aggregates of minimally conscious intrinsic entities residing at a lower scale. The latter, in turn, were formed from intrinsic entities previously existing at a still smaller scale, ultimately down to the level of elementary particles, and back to the very beginnings of our universe. If this is correct, then IIT entails an important modification of the received view about the evolutionary origin of consciousness. Although IIT may grant that consciousness, as we know it (as biological consciousness), only appeared when suitable nervous systems evolved, the theory is also committed to claim that other forms of subjective experience–albeit minimal–existed in non-biological entities since the very beginnings of our universe. Consequently, IIT also presents a radical departure from the traditional view in astrophysics and cosmology, which holds that the elementary particles created in the first moments after the Big Bang can be fully described by physics alone, and thus are unrelated to consciousness.

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

In the remainder of the article, we address a potential way in which IIT might be able to avoid the implication of a *primordial ontological dust* or "Big Bang consciousness" to account for the origin of consciousness. The motivation underlying this strategy is twofold:

on the one hand, it would align with the received view in agreeing that consciousness originated only with suitably evolved nervous systems; on the other hand, it would fully align with astrophysics in agreeing that consciousness is unrelated to the universe's fundamental ingredients formed after the Big Bang. This potential reply is inferred from Christof Koch's latest book (2024).

As shown by our analysis, IIT's implication of a primordial ontological dust ultimately follows from the theory's ″Great Divide of Being", combined with scientific knowledge regarding biological evolution, prebiotic chemical transformations, and physical cosmology. Without the notion of (primordial) ontological dust, 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.

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). But how 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, we ask, how could the Sun exist and 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 existed and shined, 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* instead of *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 is claiming that without consciousness there is only a state of undifferentiated, formless, meaningless reality; not the structured, coherent, meaningful world that we experience. Although Koch's notion of an unstructured state of existence may be difficult to grasp, if plausible, it would be certainly better than nothing (non-existence). Let us unpack

Koch’s suggestion in more detail, keeping in mind that our aim here is simply to interpret Koch’s position to the best of our ability, and leave our analysis and criticisms for the next section.

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. perceived leaves, branches, trees, etc.) that are interrelated to each other 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, according to which the mind actively structures the raw, formless sensory data through categories and concepts, thereby organizing experience into coherent objects like trees and forests. In Kant’s philosophy, objects as we know them depend on the faculties of the human mind, implying that without a conscious subject, there are no meaningful distinctions in nature, 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*, for instance, 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 significance of subjective experience (intrinsic existence) and its objects (extrinsic existence).

In sum, 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*. Hence, Koch's potential answer to the origin of consciousness problem might be the following. First, recall that Koch (2024) adheres to what we 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`s may answer that consciousness indeed evolved from non-conscious predecessors, but this doesn't entail that consciousness originated from nothing, it evolved from previously nonconscious, undifferentiated, formless stuff (which is arguably better than nothing!). We call Koch's suggestion the *formless stuff hypothesis.*

Now, this "formless stuff" can be interpreted in two different ways. According to a first interpretation, reality beyond our consciousness, and in particular, the Earth and its inhabitants before the "first flickering of phenomenal light [that] likely appeared during the Cambrian explosion" (Koch 2024, p. 105), existed as formless stuff in the sense of being reductively constituted by ontological dust of minimally conscious, intrinsic micro-entities like φ_s-maximal hydrocarbon molecules, or individual atoms. In other words, non-conscious living organisms inhabiting the Earth at a pre-consciousness, Precambrian era, would not be "nothing" but instead, non-φ_s-maximal, and thus ontologically reducible, collections of minimally conscious entities intrinsically existing at a lower scale of reality (e.g. the atomic). In this case, the *formless stuff hypothesis* turns out to be identical to what we already discussed as option (iii): CONSCIOUSNESS FROM ONTOLOGICAL DUST (section 3.3.). Moreover, we saw that indeed, all the alternatives discussed from section 3.1 to section 3.3

led to the notion of an ontological dust, and in particular, to a *primordial ontological dust* constituted by indivisible elementary particles created the first instances after the Big Bang (section 3.4.). Hence, if Koch's formless stuff hypothesis is interpreted according to this first meaning, then it also converges into the hypothesis of "Big Bang consciousness" as the theoretical endpoint to account for the origin of subjective experience.

According to a second reading of Koch's *formless stuff hypothesis*, non-conscious living organisms inhabiting our planet before the phylogenetic origin of consciousness, would not be "nothing" but instead, an undifferentiated and unstructured existence that is completely independent of consciousness. This would avoid the implication of "Big Bang consciousness" to account for the first appearance of experience. Hence, it would avoid the dramatic modifications to both the received view about the evolutionary origin of consciousness, and to the traditional picture derived from cosmology and astrophysics about the non-conscious nature of the first elementary particles.

According to this interpretation of Koch's formless stuff, 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 we may call "formless existence", that in contrast to extrinsic existence, is not grounded on the intrinsic existence of consciousness. Thus, according to this interpretation, there could be something instead of nothing in a universe completely devoid of even minimally conscious intrinsic micro-entities.

Although we find the appeal of the formless stuff hypothesis according to this second interpretation, in the following we will argue against it. More specifically, we will argue both against the plausibility of a third type of existence beyond the intrinsic/extrinsic division, as well as against the plausibility of stating that in the absence of consciousness, reality is just an undifferentiated, unstructured ontological chaos.

4.1. Refutation of the "formless stuff" hypothesis (in its second interpretation)

First, the very suggestion that besides intrinsic and extrinsic existence there is a third type of existence, namely "formless stuff", is theoretically very unstable. According to the most straightforward reading of IIT, the 'Great Divide of Being' covers all of existence—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[9].

Second, the very suggestion that without consciousness there is only formless, undifferentiated stuff, is deeply problematic for considerations pertaining to both our scientific worldview and common sense. We can see this by recalling Koch's own example: "Think of the early Earth… the planet was lifeless… Despite the sun shining, Earth's surface

[9] Thus, the formless stuff hypothesis interpreted this way, by introducing this third type of existence, undermines the very ontological consistency that IIT aims to offer as a metaphysical foundation for a comprehensive view of nature and the place of consciousness.

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, our scientific view of the non-conscious world is not of undifferentiated chaos. And this is precisely what Koch suggests when he writes that before the origin of consciousness on the early Earth the sun still shined: there was some significant, intelligible order, as inferred by planetary science, paleoclimatology, geophysics, etc. 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… during the Cambrian explosion" (Koch 2024, p. 105), even if there was nobody that 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 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 of ignoring the role of consciousness in the creation of scientific knowledge. This omission leads to a fragmented understanding of reality, where 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 the inclusion of consciousness as an essential component in understanding and constructing scientific knowledge, they do not go as far as claiming that the complexity, structure and coherence of physical reality itself is totally 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 doesn't 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 agree with Kant's fundamental 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 doesn't 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 doesn't necessarily follow from the former. Koch, in contrast, 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. It incorrectly extends an epistemological claim (about our conceptual understanding, categorization and perception of trees and forests), into an ontological claim (about the actual properties and existence of trees and forests). On the contrary, 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.

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. By acknowledging the reality of intricate interactions between non-conscious entities, at the levels of both cosmological and evolutionary history, we can better understand how consciousness could naturally evolve 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 portraits the appearance of consciousness as a sudden, inexplicable leap from something fundamentally different and almost unintelligible, contradicting the gradual change expected in evolutionary explanations.

5. Big Bang consciousness or rejection of the Great Divide?

In sum, IIT seems to be pushed to claim one of these two things. First, the theory may retain its ontological commitment to the consciousness-grounded intrinsic/extrinsic division as exhausting all of reality, and posit that the non-conscious ancestors of the first sentient creatures were reducible collections of ontological dust. In this case, given our previous argumentation, these aggregates of ontological dust should not be considered unstructured, "formless stuff". Rather, they exist between chaos and maximal integration. These aggregates should not be completely "formless", because in that case they will be unable to provide intelligible explanations for the phylogenetic origin of experience. But at the same time, they cannot be too organized and integrated either, because their system-φ must be lesser than the

φ_S of their constituent units of ontological dust. As already stated, this path ultimately leads to the necessity of positing a primordial ontological dust or "Big Bang consciousness".

Alternatively, IIT could claim that before consciousness, there was something instead of nothing, but this "something" was not "formless stuff", but already complexly organized, lawful, and differentiated, and completely independent from any consciousness. In this case, realism about a world before, and beyond, consciousness is vindicated, and the "received view" that consciousness phylogenetically originated from non-conscious living ancestors can be easily accommodated. The cost for IIT is nonetheless to abandon its "Great Divide of Being" as the ultimate, exhaustive division of all existing entities, and claim that there is genuine, structured existence wholly independent from consciousness. In other words, IIT would be forced to revise its so-called "principle of true existence", according to which "only phenomenal existence is true existence" (Cea et al., 2023, p. 4), in favor of a view that acknowledges that true existence may be dissociated from consciousness (Cea et al., 2024a).

6. Conclusions and future work

In this paper, we have presented what we term the "origin of consciousness problem" for Integrated Information Theory (IIT). The problem stems from IIT's ontological commitment that only conscious entities truly exist, while nonconscious entities might exist only from the perspective of other conscious entities, leaving a hypothetical pre-consciousness stage of life on Earth in a state of non-existence. This requires a substantial modification of the widely accepted scientific view of the evolution of consciousness, what we call the received view: that conscious life phylogenetically originated from non-conscious ancestors. If IIT holds that without consciousness there is neither intrinsic nor extrinsic existence, it faces the paradox of

claiming that consciousness originated from nothing, a notion that seems untenable both philosophically and scientifically.

We explored several potential responses IIT could offer to address this problem, including the suggestions that consciousness and life co-evolved at some point in Earth's history, or that sentient organisms evolved from non-sentient aggregates of ontological dust. Our analysis indicated that each of these responses ultimately leads, however, to the necessity of positing a form of "Big Bang consciousness", what we referred to more technically as a *primordial ontological dust*—a collection of minimally conscious (i.e., intrinsically existing) elementary particles formed at the beginning of the universe. This primordial ontological dust seems to be a theoretical endpoint for IIT to resolve the problem of the origin of consciousness. Importantly, if correct, IIT's answer would challenge the traditional view that consciousness first arose with the evolution of nervous systems, and the physical-cosmological belief that the universe's early elements are completely characterized by standard physical properties (e.g. charge, mass, spin, etc.) and hence unrelated to consciousness.

Koch's potential alternative to this "Big Bang consciousness", namely the *formless stuff hypothesis*, was also examined. It embraces the spirit of the received view, according to which consciousness evolved from non-conscious processes in living ancestors, and suggests that prior to the origin of consciousness on Earth there was something instead of nothing, an unstructured "formless stuff". We saw that this reply can be interpreted in two ways.

First, formless stuff can be seen as aggregation of ontological dust, in which case Koch's hypothesis ends up needing to posit a *primordial* ontological dust too. We argued that for this interpretation to work, these dust-aggregates cannot be either the unstructured ontological

chaos suggested by Koch, or maximally integrated wholes that exist beyond their constituents. In this case, the "received view" is only partially respected, because it agrees with the notion that consciousness "as we know it", namely, as the subjective experience of a subset of living organisms (Birch, 2024; Godfrey-Smith, 2024), could indeed have originated with the evolution of nervous systems. But at the same time, it posits that strictly speaking, consciousness pre-dated this, albeit in a very minimal form in maximal-φ_s systems probably at the molecular, atomic or subatomic levels. Even if the scale of the ontological dust constituting Precambrian, non-conscious organisms is not the elementary one, our analysis has shown that it ultimately depends as well on the first appearance of a *primordial ontological dust* or "Big Bang consciousness".

According to the second interpretation, formless stuff could be interpreted as a third type of existence. However, this proposal faces significant challenges. It neither fits with IIT's intrinsic-extrinsic distinction, nor aligns with a coherent scientific worldview. Moreover, it seems to conflate the epistemic and phenomenological role of consciousness in enabling our knowledge of the world structure with the ontological claim that there is no such world structure beyond the gaze of subjective experience. All these challenges make Koch's formless stuff–interpreted as a third type of existence not grounded in consciousness–an implausible solution to the origin of consciousness problem.

Whether the idea of a *primordial ontological dust* is in itself plausible remains an open question. However, there are compelling reasons to doubt this. First, the link between consciousness and life is worth investigating. All known conscious entities are living organisms (Birch, 2024; Godfrey-Smith, 2024; Thompson, 2007, 2015), and there are substantial reasons to consider a deep connection between being alive and being conscious (Cea, 2023b, 2023a; Cea & Martínez-Pernía, 2023; Damasio & Damasio, 2024;

Fuchs, 2018; A. Seth, 2021, 2024; A. K. Seth & Tsakiris, 2018; Thompson, 2011, 2022). This connection raises doubts about the plausibility of minimally conscious elementary particles that are non-living, as IIT's ontological dust would suggest.

Second, the notion of a primordial ontological dust requires the empirical possibility of what may be called *primordial* monads, i.e., minimally conscious, indivisible, elementary particles of intrinsic existence created at the very beginnings of our universe. But there are both formal-mathematical and conceptual challenges to the idea of minimally conscious fundamental particles. The very concept of a minimally conscious monad—an indivisible, elementary conscious entity (Hendren et al., 2024c)—appears to conflict with the formalism of IIT itself, which requires that valid partitions (i.e., divisions) to a candidate system are achievable, such that the integration postulate can be applied (and consequently the exclusion and composition postulates). In future work we argue that it is formally-mathematically inconsistent to attribute (even minimal) consciousness, and thus intrinsic existence, to such fundamental entities, precisely because they are partless and indivisible (Cea et al., 2024c). Thus, if IIT wants to preserve its Great Divide of Being and related principle of true existence ("only phenomenal existence is true existence") (Cea et al., 2023), the theory might need to reject the view that minimally conscious elementary particles are indivisible and not constituted by even smaller entities as parts, challenging the very notion of fundamental monads, and the physical notion of elementary particles as fundamental ingredients of the universe that are not further composed of smaller constituent parts (Barrett, 2014).

In light of these and other challenges (Cea et al., 2024b; Signorelli et al., 2023), IIT may instead benefit from revising its ontological assumptions. Specifically, future work should analyze IIT's notion of "true existence", and the plausibility of extending it beyond the

intrinsic existence of consciousness (Cea et al., 2024a). Perhaps all causally powerful entities—whether conscious or not—deserve to be granted some form of genuine existence, rejecting IIT's "Great Divide of Being" as an exhaustive ontological division between all existing entities. Such a revision would require a careful examination of the taxonomy of entities that IIT's formalism allows, determining which entities, conscious or not, legitimately exist, based on a re-assessment of the relationship between consciousness, causality and existence (Cea et al., 2024a).

In conclusion, the origin of consciousness problem forces IIT to confront both its assumptions about the nature of existence and its formal metaphysics. It currently entails that consciousness really originated in the beginnings of our universe in the form of a primordial ontological dust or "Big Bang consciousness". However, this hypothesis challenges what we called the "received view" about the phylogenetic origin of consciousness, as well as the standard physical understanding of the nature of the first elementary particles that existed, which is not commonly related to consciousness. Thus, future work must address the plausibility of the notion of a primordial ontological dust and its constituent monads. Also, we should consider whether alternative metaphysical frameworks may be compatible with IIT's formalism, such as forms of emergentism (Cea, 2021; Negro, 2022b) or conscious agents (Prentner & Hoffman, 2024). These may include different ontological levels of causally powerful entities as truly existing despite not being necessarily conscious, thus potentially offering a more coherent and scientifically grounded account for the origin of consciousness.

Author Contributions

IC: Conceptualization, Supervision, Visualization, Writing – original draft, Writing – review & editing. NN: Conceptualization, Writing – review & editing. CMS: Conceptualization, Writing – review & editing.

Statements and declarations:

Ethical considerations

Not applicable

Consent to participate

Not applicable

Consent for publication

Not applicable

Declaration of conflicting interest

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding statement

IC acknowledges the financial support of the CIIC-UCT. NN wishes to acknowledge the financial support of the Azrieli Foundation. CMS acknowledges the financial support by FNRS, grant Embodied-Time - 40011405, and Carlsberg Foundation, CPAI grant CF22-1432

Data availability

Not applicable

References

Albantakis, L. (2023). *Reviewer 2 Report on Cea et al., 2023 "The fundamental tension in IIT" 4.0's realist idealism'* . Entropy. https://www.mdpi.com/1099-4300/25/10/1453/review_report

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., Barbosa, L., Findlay, G., Grasso, M., Haun, A. M., Marshall, W., Mayner, W. G., Zaeemzadeh, A., Boly, M., Juel, B. E., Sasai, S., Fujii, K., David, I., Hendren, J., Lang, J. P., & Tononi, G. (2022). *Integrated information theory (IIT) 4.0: Formulating the properties of phenomenal existence in physical terms*.

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.

Barrett, A. B. (2014). An integration of integrated information theory with fundamental physics. *Frontiers in Psychology*, *5*, 63.

Binet, A. (1889). *The psychic life of micro-organisms*. Open court publishing Company.

Birch, J. (2024). *The Edge of Sentience: Risk and Precaution in Humans, Other Animals, and AI*.

Block, N. (2007). On a Confusion about a Function of Consciousness. In *Consciousness, Function, and Representation. Collected Papers, Volume 1* (pp. 159–214). MIT Press. https://doi.org/10.7551/mitpress/2111.003.0012

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/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*, 10597123231184652. https://doi.org/https://doi.org/10.1177/10597123231184651

Cea, I. (2023b). The Somatic Roots of Affect: Toward a Body-Centered Education. In P. Fossa & C. Cortés-Rivera (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/https://doi.org/10.53765/20512201.30.3.007

Cea, I., Negro, N., & Signorelli, C. M. (2023). The Fundamental Tension in Integrated Information Theory 4.0's Realist Idealism. *Entropy 2023, Vol. 25, Page 1453*, *25*(10), 1453. https://doi.org/10.3390/E25101453

Cea, I., Negro, N., & Signorelli, C. M. (2024a). *How to be an Integrated Information Theorist without losing your body* (In Preparation).

Cea, I., Negro, N., & Signorelli, C. M. (2024b). Only Consciousness Truly Exists? Two Problems for IIT 4.0's ontology. *Frontiers in Psychology*.

Cea, I., Negro, N., & Signorelli, C. M. (2024c). *Why Phi-Monads cannot exist: IIT 4.0 and the formal impossibility of indivisible units of consciousness*.

Chalmers, D. (1996). *The Conscious Mind: In Search of a Fundamental Theory*. Oxford University Press.

Chalmers, D. (2021). Idealism and the mind-body problem. In *The routledge Handbook of idealism and immaterialism* (pp. 591–613). Routledge.

Chis-Ciure, R. (2022). The transcendental deduction of Integrated Information Theory: connecting the axioms, postulates, and identity through categories. *Synthese*, *200*(3), 1–27.

Consortium, C., Ferrante, O., Gorska-Klimowska, U., Henin, S., Hirschhorn, R., Khalaf, A., Lepauvre, A., Liu, L., Richter, D., & Vidal, Y. (2023). An adversarial collaboration to critically evaluate theories of consciousness. *BioRxiv*, 2023–2026.

Damasio, A., & Damasio, H. (2023). Feelings Are the Source of Consciousness. *Neural Computation*, *35*(3), 277–286. https://doi.org/10.1162/neco_a_01521

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

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., P. Lang, J., M. Haun, A., Albantakis, 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*, 667.

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. (1998). *Complexity and the Function of Mind in Nature*. Cambridge University Press.

Godfrey-Smith, P. (2016). *Other minds: The octopus, the sea, and the deep origins of consciousness*. Farrar, Straus and Giroux.

Godfrey-Smith, P. (2024). *Living on Earth: forests, corals, consciousness, and the making of the world*.

Graziano, M. S. A. (2017). The attention schema theory: A foundation for engineering artificial consciousness. *Frontiers in Robotics and AI*, *4*, 60.

Graziano, M. S. A. (2019). *Rethinking consciousness: a scientific theory of subjective experience*. WW Norton & Company.

Haeckel, E. (1892). Our monism. The principles of a consistent, unitary world-view. *The Monist*, 481–486.

Hendren, J., Grasso, M., Juel, B. E., & Tononi, G. (2024a). *IIT Wiki - How does intrinsic in IIT relate to the notion as used in philosophy (e.g. Lewis)?* Https://Www.Iit.Wiki/Faqs/Philosophy#h.Ky7wm6ipobmt.

Hendren, J., Grasso, M., Juel, B. E., & Tononi, G. (2024b). *IIT Wiki - Identity as a good explanation*. Https://Www.Iit.Wiki/Identity#h.Jz0i82fsjg65.

Hendren, J., Grasso, M., Juel, B. E., & Tononi, G. (2024c). *IIT Wiki - Monad*. Https://Www.Iit.Wiki/Glossary#h.Otx4utf9u8sn.

Hoel, E. P., Albantakis, L., Marshall, W., & Tononi, G. (2016). Can the macro beat the micro? Integrated information across spatiotemporal scales. *Neuroscience of Consciousness*, *2016*(1), niw012. https://doi.org/10.1093/nc/niw012

Hordijk, W., Hein, J., & Steel, M. (2010). Autocatalytic sets and the origin of life. *Entropy*, *12*(7), 1733–1742.

Hordijk, W., & Steel, M. (2017). Chasing the tail: The emergence of autocatalytic networks. *Biosystems*, *152*, 1–10.

Humphrey, N. (2022). *Sentience: The invention of consciousness*. Oxford University Press.

Jain, S., & Krishna, S. (1998). Autocatalytic sets and the growth of complexity in an evolutionary model. *Physical Review Letters*, *81*(25), 5684.

Jennings, H. S. (1931). *Behavior of the lower organisms*. Columbia University Press.

Kauffman, S. A. (1993). *The Origins of Order: Self-Organization and Selection in Evolution*. Oxford 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.

Margulis, L., & Sagan, D. (2000). *What is life?* Univ of California Press.

Marshall, W., Albantakis, L., & Tononi, G. (2018). Black-boxing and cause-effect power. *PLoS Computational Biology*, *14*(4), e1006114.

Marshall, W., Kim, H., Walker, S. I., Tononi, G., & Albantakis, L. (2017). How causal analysis can reveal autonomy in models of biological systems. *Philosophical Transactions of the Royal Society a: Mathematical, Physical and Engineering Sciences*, *375*(2109), 20160358.

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 E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2021). Metaphysics Research Lab, Stanford University.

Nagel, T. (1974). What is it like to be a bat? *Philosophical Review*, *83*(October), 435–450.

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.

Oizumi, M., Albantakis, L., & Tononi, G. (2014). From the Phenomenology to the Mechanisms of Consciousness: Integrated Information Theory 3.0. *PLoS Computational Biology*. https://doi.org/10.1371/journal.pcbi.1003588

Prentner, R., & Hoffman, D. D. (2024). Interfacing consciousness. *Frontiers in Psychology*, *15*, 1429376.

Reber, A. S., Baluska, F., & Miller, W. (2023). *The sentient cell: the cellular foundations of consciousness*. Oxford University Press.

Seth, A. (2021). *Being you: A new science of consciousness*. Penguin.

Seth, A. (2024). Conscious artificial intelligence and biological naturalism. *PsyArXiv*. https://doi.org/https://doi.org/10.31234/osf.io/tz6an

Seth, A. K., & Bayne, T. (2022). Theories of consciousness. *Nature Reviews Neuroscience*, *23*, 439–452.

Seth, A. K., & Tsakiris, M. (2018). Being a Beast Machine: The Somatic Basis of Selfhood. *Trends in Cognitive Sciences*, *22*(11), 969–981. https://doi.org/10.1016/j.tics.2018.08.008

Signorelli, C. M., Cea, I., & Prentner, R. (2023). We need to explain subjective experience, but its explanation may not be mechanistic. *Psyarxiv*.

Signorelli, C. M., Szczotka, J., & Prentner, R. (2021). Explanatory profiles of models of consciousness-towards a systematic classification. *Neuroscience of Consciousness*, *2021*(2), niab021.

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. (2011). Précis of mind in life: Biology, phenomenology, and the sciences of mind. In *Journal of Consciousness Studies*.

Thompson, E. (2015). *Waking, dreaming, being: Self and consciousness in neuroscience, meditation, and philosophy*. Columbia University Press.

Thompson, E. (2022). Could All Life Be Sentient? *Journal of Consciousness Studies*, *29*(3–4), 229–265.

Tononi, G. (2017). Integrated information theory of consciousness: Some ontological considerations. In S. Schneider & M. Velmans (Eds.), *The Blackwell companion to consciousness (2nd edition)* (pp. 621–633). Wiley Online Library. https://doi.org/https://onlinelibrary.wiley.com/doi/10.1002/9781119132363.ch44

Tononi, G., Albantakis, L., Boly, M., Cirelli, C., & Koch, C. (2022). Only what exists can cause: An intrinsic view of free will. *ArXiv Preprint ArXiv:2206.02069*.

Tononi, G., & Koch, C. (2015). Consciousness: Here, there and everywhere? In *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.

Weiss, M. C., Sousa, F. L., Mrnjavac, N., Neukirchen, S., Roettger, M., Nelson-Sathi, S., & Martin, W. F. (2016). The physiology and habitat of the last universal common ancestor. *Nature Microbiology*, *1*(9), 1–8.