

CHAPTER 6

THE COGNITIVE PHENOMENOLOGY ARGUMENT FOR DISEMBODIED AI CONSCIOUSNESS

Cody Turner
University of Connecticut., USA

Abstract: In this chapter, I shall offer two novel arguments for what I call *strong primitivism* about cognitive phenomenology, the thesis that there exists a phenomenology of cognition that is neither reducible to, nor dependent upon, sensory phenomenology. I will then contend that strong primitivism implies that phenomenal consciousness does not require sensory processing. This latter contention has implications for the philosophy of artificial intelligence. For if sensory processing is not a necessary condition for phenomenal consciousness, then it plausibly follows that AI consciousness (assuming that it is possible) does not require embodiment. The overarching goal of this chapter is to show how different topics in the analytic philosophy of mind can be brought to bear on an important issue in the philosophy of artificial intelligence.

The central argument considered here is called ‘the Cognitive Phenomenology Argument for Disembodied AI Consciousness’. It is as follows:

- P1.** Strong primitivism about cognitive phenomenology is true;
- P2.** If strong primitivism about cognitive phenomenology is true, then phenomenal consciousness does not require sensory processing;
- C1.** Therefore, phenomenal consciousness does not require sensory processing;
- P3.** If phenomenal consciousness does not require sensory processing, then AI consciousness (assuming that it is possible) does not require embodiment;
- C2.** Therefore, AI consciousness (assuming that it is possible) does not require embodiment.

As indicated, I first focus on P1 by presenting two arguments for strong primitivism. The first argument is a phenomenological argument based on the possibility of what some Buddhists and mystics call *pure consciousness*. The second argument draws from the *higher-order thought theory of consciousness*. Importantly, neither of these arguments is intended (in this chapter at least) to definitively demonstrate the veracity of P1. The aim of both arguments is to show that strong primitivism is plausible given certain theoretical assumptions.¹ My goal here, in other words, is merely to establish the following two conditionals:

¹ I am strictly speaking agnostic both with respect to whether pure consciousness is possible, and with respect to whether the higher-order thought theory of consciousness is true.

(CON1) If pure consciousness is possible, then strong primitivism about cognitive phenomenology is true.

(CON2) If the higher-order thought theory of consciousness is true, then strong primitivism about cognitive phenomenology is true.

After arguing for CON 1 and CON 2, I will proceed to justify P2 and P3 of the cognitive phenomenology argument for disembodied AI consciousness. Before turning to the argument though, I will overview the cognitive phenomenology debate.

1. The Cognitive Phenomenology Debate

Until the 1980s, it was mainstream in analytic philosophy of mind to view phenomenality and intentionality as two separate aspects of the mental. This view, which Terrence Horgan and John Tienson (2002) call *separatism*, holds that phenomenal states and intentional states are metaphysically independent of one another, meaning that neither kind of state arises from the other and that a system can exhibit one and not the other, at least in principle. Separatism is traditionally motivated by the conviction that while occurrent *sensory* mental states (e.g., the state you are in when seeing the color red) possess phenomenal character but lack intentional content, occurrent *cognitive* mental states (e.g., the state you are in when believing that Hartford is the capital of Connecticut) possess intentional content but lack phenomenal character.

Cognitive phenomenology (CP for short) is a concept in the philosophy of mind that functions to undermine separatism. Proponents of CP assert that occurrent cognitive states are both intentionally directed and (in some sense at least) phenomenally conscious. According to these proponents, there is an experiential character associated with thinking, where ‘thinking’ is understood to encompass a wide range of propositional attitudes, such as believing, doubting, wondering, desiring, entertaining, etc.

One can distinguish in conceptual space between four different views of cognitive phenomenology:

1. **Eliminativism about CP:** Cognitive phenomenology does not exist.
2. **Reductionism about CP:** Cognitive phenomenology is reducible to sensory phenomenology.
3. **Weak Primitivism about CP:** Cognitive phenomenology is irreducible to sensory phenomenology, but dependent upon sensory phenomenology.
4. **Strong Primitivism about CP:** Cognitive phenomenology is irreducible to, and independent of, sensory phenomenology.

Before elucidating these four positions, it is important to highlight that cognitive phenomenology only pertains to *occurrent* cognitive states. *Dispositional* cognitive states, such as all of my beliefs that I am not currently entertaining, are not phenomenally conscious. For there is nothing

it is like to have a belief that is not being presently tokened. With this being said, let us turn towards the four positions on CP.

Eliminativists about CP concede that occurrent cognitive activities are *access conscious*, but they deny that such activities are *phenomenally conscious*. A mental state is access conscious if and only if it is available for verbal report, rational inference, and/or the deliberate control of behavior (Block, 1995). While eliminativists acknowledge that occurrent desires, beliefs, wonderings, etc. are available for these uses, they reject the notion that there is something it is like to be in these cognitive states.² By contrast, reductionists grant that cognition involves phenomenology, but uphold that cognitive phenomenology can be explained in terms of sensory phenomenology that typically accompanies thinking, such as mental imagery and inner speech.

The debate between eliminativism and reductionism about CP may ultimately just be a verbal dispute. The same cannot be said with respect to the debate between reductionism and primitivism. Primitivists agree with reductionists that various forms of sensory phenomenology, such as auditory imagery (i.e., inner speech) and visual imagery (i.e., mental pictures), often accompany our thoughts. Unlike reductionists, however, primitivists renounce the claim that this concomitant sensory phenomenology *exhausts* the experiential character of occurrent cognition. Consider the thought <buses are yellow>. Each tokening of this thought may be, and often is, connected with a distinct sensory phenomenal character that represents it. For example, in one instance the thought may be represented by a mental image of a yellow bus whereas in another it may be represented by an inner speech sentence that says ‘BUSES ARE YELLOW.’ Primitivists aver that although these two tokenings of the same thought differ with respect to their sensory character, there is nevertheless a common ‘what-it’s-likeness’ that runs through them. This shared experiential character, according to the primitivist, denotes a *sui generis* cognitive phenomenology, one that cannot be reduced to any conjunction of auditory and/or visual imagery.

The distinction between *weak* primitivism and *strong* primitivism about CP is rarely made in the literature on cognitive phenomenology (Levine, 2011, is an exception). This distinction is motivated by the insight that phenomenal irreducibility does not entail phenomenal independence. X phenomenally depends upon Y iff X cannot be instantiated in the absence of Y. Weak primitivism holds that while cognitive phenomenology is irreducible to sensory phenomenology, cognitive phenomenology cannot be instantiated in the absence of sensory phenomenology. As Levine states, on weak primitivism, “it’s still the case that the only way to phenomenally experience a cognitive content is through its effect on some sensory presentation” (Levine, 2011: 112). Strong primitivism, on the other hand, states that cognitive phenomenology is both phenomenally irreducible to, and phenomenally independent of, sensory phenomenology.

2. The ‘Pure Consciousness’ Argument for Strong Primitivism

Some of the most popular arguments for primitivism about CP are phenomenological in nature. These phenomenological arguments often take the form of *phenomenal contrast cases*

² See Carruthers and Veillet (2011) for an argument in support of eliminativism.

(Chudnoff, 2015). For example, the phenomenology of an English speaker is contrasted with the phenomenology of a person who does not speak English. We are asked to imagine for the sake of argument that both the English speaker and the non-English speaker hear the same English sentence uttered by the same person at the same exact location. There is in this imagining, so the argument goes, an overall phenomenal difference between the two speakers despite the fact that their respective sensory phenomenologies are identical. This phenomenological insight is taken to suggest that there must be an irreducible cognitive phenomenology associated with understanding language.

Levine (2011) correctly notes that most of the existing phenomenological arguments for primitivism about CP at best support weak primitivism (he calls weak primitivism ‘impure CP’ and strong primitivism ‘pure CP’):³

When it comes to the distinction between pure and impure CP, it seems obvious that this argument [the phenomenological contrast argument for primitivism] can’t support more than impure CP. After all, all the examples involve ways of experiencing visual appearances, sound streams, and the like. The argument is addressing someone who doesn’t find anything but sensory phenomenology in her conscious life, and it pushes her to notice distinctions among these sensory appearances that can only be accounted for by appeal to cognitive penetration. Fair enough. Still, this doesn’t get us anything like pure CP. It’s still possible, for all this argument demonstrates, that the only way for a cognitive content to make itself appear to a conscious subject is through affecting the way some sensory manifold appears. That’s impure CP (Levine, 2011: 115-116).

This chapter offers a phenomenological argument for strong primitivism that is based upon the possibility of ‘pure consciousness’. Pure consciousness is a phenomenal state that is supposed to be achievable via mediation.

It is claimed that, during instances of pure consciousness, the intentional structure of experience disappears so as to leave nothing but an open phenomenal field. David Woodruff Smith (2011) describes the state of pure consciousness thusly: “In sustained meditation the sense of object disappears, i.e., the intentional content or object-specifying sense disappears, and the sense of self or subject disappears. That is to say, the usual intentional structure of subject-object-or subject-act-content-object-dissolves in such a meditative state. There is then just this flowing phase of consciousness” (Smith, 2011: 481). The ‘pure consciousness’ argument for strong primitivism can be represented in the following manner:

PP1. Pure consciousness is possible;

PP2. If pure consciousness is possible, then strong primitivism about cognitive phenomenology is true;

C1. Therefore, strong primitivism about cognitive phenomenology is true.

PP1 is supported by the phenomenology of deep meditative experience. One could push back against PP1 by, for example, arguing against the common wisdom that meditation is a reliable epistemic guide to one’s own phenomenal states. One could also reject PP1 by endorsing Franz

³ A notable exception is Chudnoff (2015). There he presents a phenomenal contrast case that does in fact support strong primitivism about CP.

Brentano's (1874) thesis that intentionality is the mark of the mental. If Brentano is correct in stating that mental states are necessarily intentional, then it follows that pure conscious mental states are impossible, for such states are reportedly devoid of intentionality. Steven Katz (1978) argues against the possibility of pure consciousness along these lines. According to Katz, states of pure consciousness are unachievable because all conscious experiences (including mystical experiences) are mediated by the presence of some conceptual schema: "The experience itself as well as the form in which it is reported is shaped by concepts which the mystic brings to, and which shape, his experience" (Katz, 1978). I take the plausibility of PP1 for granted in this chapter, and focus on PP2.

PP2 is *prima facie* tenable given that pure consciousness by stipulation exists in the absence of sensory phenomenology. I see two ways that one could go about objecting to PP2. The first is to *deny that pure consciousness involves phenomenal consciousness*. Perhaps the alleged 'emptiness' of states of pure consciousness is due to the fact that such states lack phenomenal character. If pure conscious states are devoid of phenomenal character, then pure consciousness is not a kind of cognitive phenomenology, contra to what PP2 suggests. Proponents of pure consciousness are very explicit, however, that there is a *what-it's-likeness* associated with being in a pure conscious state. As Smith emphasizes: "This meditative experience is *phenomenal*: it is appearing, "phenomenally", in consciousness, and what it is like is simply one of being conscious, as if awaiting what else may or may not appear in consciousness" (Smith, 2011: 481). This quote from Smith indicates that phenomenality is built into the very concept of pure consciousness. If this is the case, then the present objection (that pure consciousness is non-phenomenal) seems to be at its core an objection to PP1, rather than to PP2 as previously supposed.

A different way to go about objecting to PP2 is to *concede that pure consciousness is phenomenal, but deny that pure consciousness is cognitive*. According to this objection, the phenomenology of pure consciousness involves neither sensory phenomenology nor cognitive phenomenology. A phenomenally conscious mental state X, so the objection goes, qualifies as cognitive if and only if X is characterized by a phenomenally conscious cognitive attitude and/or phenomenally conscious cognitive contents. The problem is that states of pure consciousness involve neither cognitive attitudes nor cognitive contents. This is because such states are by definition synonymous with the disappearance of the intentional structure of experience. The phenomenology of pure consciousness is therefore not a kind of cognitive phenomenology. Or so the objection goes at least.

The claim that pure consciousness is cognitive in nature can, I propose, be reconciled with the mystics' claim that the state of pure consciousness reveals no discernible intentional structure. The idea I have in mind is this. Pure consciousness is an intentional mental state, but it is an intentional state that has as its object the concept of nothingness, or perhaps the concept of non-intentionality itself. Put differently, the idea is that pure consciousness is an intentional state that represents the state of non-representation. Construing pure consciousness in this light allows one to maintain that pure consciousness is a type of cognitive phenomenology while acknowledging that states of pure consciousness seem to be devoid of intentionality from a first-person point of view.

This proposed analysis of pure consciousness is similar in many respects to how various philosophers of mind analyze the phenomenology of *moods*. Moods, such as elation and anxiety, are sometimes thought to be non-intentional, undirected mental states. Advocates of Brentano's thesis of course reject this conception of moods as non-intentional. Said advocates instead claim that "while moods *appear* not to be directed at anything, upon closer examination, it turns out that they are in fact directed at special kinds of objects: bodily states or unusual external objects, such as the world as a whole, indeterminate intentional objects, or frequently changing objects (see Goldie, 2000; 2002; Seager, 2002; Seager and Bourget, 2007 and Tye, 2008)" (Mendelovici, 2013: 8-9). Just as some philosophers try to explain away the apparent non-intentionality of moods, I am here trying to explain away the apparent non-intentionality of pure conscious states. More specifically, in the same way that the intentional object of states of anxiety or elation might be 'the world as a whole' or 'everything', I am suggesting that the object of states of pure consciousness might be 'the absence of an intentional world', or 'nothing'. If this is right, then it is unsurprising that states of pure consciousness are (mistakenly) reported to be devoid of intentionality.

The objector to PP2 may at this juncture in the dialectic accuse me of changing the subject, or at the very least, of misunderstanding the concept of pure consciousness. The mystics' claim, the objector may insist, is not merely that pure consciousness *appears* to lack intentionality; it is rather the more robust claim that pure consciousness *as a matter of fact* lacks intentionality. My response to this retort is in effect to double down: pure consciousness is 'cognitive' precisely because pure conscious states are, despite phenomenological appearances, actually intentional.

Perhaps pure consciousness is by definition non-intentional, though. I grant that states of pure consciousness cannot properly be called 'cognitive' phenomenal states if this is the case. The fact of the matter, however, is that a phenomenological argument for strong primitivism can be given without appealing to the notion of pure consciousness. *Even if pure consciousness is not a kind of cognitive phenomenology, it is not difficult to imagine a slightly modified phenomenological argument for strong primitivism that is based upon the phenomenology of meditation.* One might argue, for example, that isolated awareness of propositional contents during 'mindfulness' meditation (or during immersion in a sensory deprivation tank) constitutes phenomenological evidence in support of strong primitivism. According to Smith, "observant self-awareness is the goal of the practice called 'mindfulness' or *vipassana* in one Buddhist tradition. In this practice, the meditative experience prepares for an experience of consciousness per se – pure awareness, as it were – and allows then for noticing particular thoughts, feelings, or sensations that may come along in the stream of consciousness" (Smith, 2011: 481). Isolated awareness of propositional contents is not pure consciousness, but it is arguably a phenomenal experience that lends credence to strong primitivism.

3. The 'Higher-Order Thought' Argument for Strong Primitivism

I now turn to the second argument for strong primitivism, which is motivated by the higher-order thought (HOT) theory of consciousness. The higher-order thought theory of consciousness, the chief proponent of which is David Rosenthal (1986, 1993, 2005), affirms that a mental state X becomes conscious when X is taken as the object of a higher-order thought. Rosenthal takes the relationship between the HOT and the first-order mental state to be one of representation (as

opposed to say, acquaintance). The argument from HOT theory to strong primitivism has the following basic structure:

- QP1.** HOT theory is true;
- QP2.** HOT theory entails strong primitivism about CP;
- C1.** Therefore, strong primitivism about CP is true.

Two important clarificatory remarks are in order with respect to the scope of this argument. First, the argument concerns only *actualist* versions of HOT theory. Actualist HOT theory is contrasted with *dispositional* HOT theory, the leading proponent of which is Peter Carruthers (2005). Dispositional HOT theory asserts that the relevant higher-order thoughts are dispositional mental states as opposed to occurrent ones. Gennaro explains that “the basic idea [behind dispositional HOT theory] is that the conscious status of an experience is due to its *availability* to higher-order thought...Thus, no actual HOT occurs” (Gennaro, 2018). Dispositional HOT theory is by stipulation not included within the purview of the HOT argument for strong primitivism.

The second clarificatory remark regarding the argument is that the argument only concerns versions of HOT theory that aim to explain *phenomenal consciousness*, or the ‘what-it’s-likeness’ of experience. HOT theory is first and foremost a theory of *state consciousness*, not phenomenal consciousness. The term ‘state consciousness’, as I am using it here, denotes the idea that a conscious mental state is a state that one is aware of being in. Thus, a mental state X of subject S possesses state consciousness iff S is aware of X. State consciousness and phenomenal consciousness are not necessarily one and the same thing, meaning that HOT theories are not necessarily theories of phenomenal consciousness.

There is considerable controversy over what the relationship is between state consciousness and phenomenal consciousness (if any). Some philosophers, such as Lycan (1996), suggest that state consciousness is a necessary condition for phenomenal consciousness. Others, such as Block (2011), believe that phenomenal consciousness outstrips state consciousness in the sense that one can be phenomenally conscious of X despite not being aware of X. Block calls this idea ‘phenomenological overflow’. I remain neutral regarding the relationship between state consciousness and phenomenal consciousness. The only point that matters for the purposes of this discussion is that the HOT argument for strong primitivism applies only to versions of HOT theory that purport to explain both state consciousness *and* phenomenal consciousness. One might immediately retort here that HOT theories are only fit (and only deployed for that matter) to explain the former type of consciousness, and that consequently the HOT argument for strong primitivism is a nonstarter.

I grant that dispositional HOT theory is ill-equipped to account for phenomenal consciousness. After all, phenomenally conscious states are by nature occurrent mental states, not dispositional ones. Rosenthal agrees: “Being disposed to have a thought about something doesn’t make one conscious of that thing, but only potentially conscious of it” (Rosenthal, 2004: 28). Fortunately, as I have already flagged, dispositional HOT theories are not included within the scope of the HOT argument for strong primitivism. The fact that dispositional theories cannot explain phenomenal consciousness does not therefore undercut the argument.

The real question is whether *actualist* HOT theories can be equipped to explain phenomenal consciousness. I will not attempt to answer this question here. I will point out though that Rosenthal's actualist HOT theory, which is the most prominent version of the view, is supposed to explain phenomenal consciousness as well as state consciousness. As Rosenthal himself states: "As many, myself included, use that phrase, there being something it's like for one to be in a state is simply its seeming subjectively that one is in that state. Indeed, Block (2011: 424) quotes me to that effect: 'What it is like for one to have a pain, in the relevant sense of that idiom, is simply what it is like for one to be conscious of having that pain'" (Rosenthal, 1997: 433).

I am agnostic as to whether Rosenthal's HOT theory is plausible. As mentioned, the only aim here is to justify QP2, the claim that (actualist) HOT theory (about phenomenal consciousness) entails strong primitivism. One might doubt QP2 on the grounds that first-order sensory states are necessary for phenomenal consciousness on HOT theory. If HOT theory holds that consciousness requires the presence of first-order sensory states, then it is presumably false that HOT theory entails strong primitivism about cognitive phenomenology. In fact, HOT theory is arguably incompatible with strong primitivism if it asserts that first-order sensory states are necessary for consciousness. For it plausibly cannot simultaneously be the case that both (a) cognitive phenomenology is irreducible to, and independent of, sensory phenomenology, and (b) phenomenal consciousness necessitates the presence of first-order sensory states.

Fortunately for the proponent of the HOT argument for strong primitivism, HOT theory is not actually committed to (b). This is because the theory upholds that the higher-order thoughts are themselves sufficient for consciousness. Rosenthal says exactly this: "And on that construal of 'what it's like,' the theory does hold that a HOT is sufficient for there to be something it's like for one to be in the state the HOT describes, even if that state doesn't occur" (Rosenthal, 1997: 433-434). Call this "the sufficiency condition":

The Sufficiency Condition: Lone higher-order thoughts are sufficient for phenomenal consciousness.

Rosenthal adopts the sufficiency condition in order to account for cases of *misrepresentation*. Where there is representation, there is also the possibility of misrepresentation. This modal truth gives rise to the following question, which the HOT theorist must answer. Namely, what happens when a higher-order thought misrepresents the content of an occurrent first-order perceptual state? Is the resulting conscious state characterized by the veridical content of the first-order state, or by the misrepresented content of the higher-order state? The former option, according to Rosenthal, is inadequate because it seems to render the higher-order state completely irrelevant with respect to the production of phenomenal consciousness. To ensure that higher-order thoughts play an essential role in generating consciousness, Rosenthal denies that first-order sensory states are required for phenomenality. According to Rosenthal, consciousness is instantiated during cases of *targetlessness*, or cases in which a higher-order thought represents there to be a first-order state that does not actually exist. Sam Coleman summarizes Rosenthal's position thusly:

How does it feel to be aware of oneself as instantiating a red London bus- is percept when no such percept obtains? Conscious mental life is the stuff of appearances, Rosenthal emphasizes: thus, since the appearance to the subject is the same whether or not she tokens a red London bus-sh percept (providing the

same HOT is in place), her experience is indistinguishable across both cases—as hallucinations can be indistinguishable from their veridical counterparts. Corresponding remarks apply to misrepresentation: HOTs govern the subjective mental appearances constitutive of the subject's stream of consciousness (Coleman, 2015: 2705).

It is the sufficiency condition that explains why HOT theory implies strong primitivism. For to say that HOTs can instantiate phenomenal consciousness in the absence of first-order sensory states is precisely to say the cognitive phenomenology does not depend upon sensory phenomenology. I am admittedly not the first person to make this argument. Richard Brown and Pete Mandik (2012) explicitly make the connection between HOT's sufficiency condition and cognitive phenomenology:

This sufficiency claim about HOTs is a key commonality between HOT theory and one of the central theses defended by cognition phenomenology proponents: A thought can give rise to there being something it's like, and does not do so in virtue of there being some co-occurring non-cognitive state (Brown & Mandik, 2012: 4).

It is worth highlighting that Brown & Mandik do not distinguish between different views of cognitive phenomenology in their paper. It can be discerned that Brown & Mandik are concerned with primitivism about cognitive phenomenology, but they do not differentiate between weak primitivism and strong primitivism. I differentiate between these two versions, and am making the specific contention that HOT theory supports strong primitivism. The opponent of the higher-order thought argument for strong primitivism might, at this stage in the dialectic, concede that the sufficiency condition is built into Rosenthal's HOT theory, but maintain that there are other actualist versions of HOT theory that do not endorse the condition.⁴

Sam Coleman (2015) defends a version of HOT theory that he claims is not committed to the sufficiency condition. Coleman calls his view the *quotational* higher-order thought (QHOT) theory of consciousness. According to QHOT theory, the relationship between the higher-order thought and the first-order state is not one of mental representation (as Rosenthal would have it), but rather one of "mental quotation." Coleman contends that mental quotation, unlike mental representation, allows for "genuine *intimacy* with our sensory states, in sensory consciousness" (Coleman, 2015: 2711).⁶ The upshot of this contention for Coleman is that first-order sensory states are necessary for phenomenal consciousness on QHOT theory. Coleman explicitly rejects the notion that isolated QHOTs are sufficient for consciousness:

What of targetlessness? A lone QHOT won't suffice for a conscious state. Here's why. QHOTs are hypothesized to supply subjective awareness. However, if a QHOT has no sensory target to embed, it could at most arouse subjective awareness of (a state of) nothing, since it altogether lacks first-order content. But subjective awareness of a state of nothing at all just isn't subjective awareness, since subjective awareness must be (intentionally) of something. An experience literally of nothing is simply no experience. Therefore an empty QHOT won't produce subjective awareness—or a conscious state. QHOTs can only make the subject aware of something if there is something (sensory content) to be aware of (Coleman 2015: 2722).

If QHOT theory is in fact devoted to the claim that first-order sensory states are necessary for consciousness. The proponent of the HOT argument for strong primitivism might respond here

⁴ Carruthers (2008) makes this exact point when discussing the sufficiency condition.

by maintaining (contra Coleman) that QHOT theory is committed to the sufficiency condition. The response that I have in mind involves *appealing to the possibility of pure consciousness*. The concept of pure consciousness, as discussed, suggests that phenomenal experience can exist in the absence of any intentional structure. If pure consciousness is possible, then Coleman is wrong in stating that “an experience literally of nothing is simply no experience” (Coleman, 2015: 2722). More to the point: there is reason to believe that targetless QHOTs do suffice for phenomenal experience given the possibility of pure consciousness.

Assume for the sake of argument, however, that Coleman is correct in believing that QHOT theory rejects the sufficiency condition. One might grant this assumption but nevertheless insist that QHOT theory entails strong primitivism about cognitive phenomenology. Specifically, the proponent of the HOT argument for strong primitivism might concur with Coleman that phenomenal experiences are necessarily *of something*, but reject Coleman’s assumption that this “something” has to be a *sensory* state. One might ask: what happens when a QHOT takes as its object a *first-order thought* (as opposed to a first-order sensory state)? Coleman presumably avers that consciousness is not instantiated in these cases, for he, to repeat, believes that phenomenal experience requires the presence of first-order sensory states on QHOT theory. An interlocutor might press Coleman though by questioning why the first-order mental states have to be sensory in nature. If QHOTs function to bestow consciousness upon first-order sensory states, then why would they not also function to bestow consciousness upon first-order cognitive states? Consistency arguably demands that first-order cognitive states be made phenomenally conscious on the theory when embedded within a QHOT. Brown & Mandik raise this point in the context of discussing HOT theory in general.⁵

If first-order thoughts are rendered phenomenally conscious when embedded in QHOTs, then QHOT theory implies strong primitivism. To repeat, this conditional remains true even if it is granted that QHOT theory rejects the sufficiency condition.

I am neutral about whether QHOT theory entails strong primitivism. Perhaps the HOT argument for strong primitivism is only tenable with respect to Rosenthal’s HOT theory. Nevertheless, the HOT argument is clearly still interesting as Rosenthal’s HOT theory is fairly prominent in the consciousness literature.

4. Premise 2: From Strong Primitivism to the Non-Necessity of Sensory Processing

Assume now, for the sake of argument, that strong primitivism about cognitive phenomenology is true. Why think that this phenomenological truth has metaphysical implications? In other words, why think that it follows from the fact that cognitive phenomenology does not depend upon sensory phenomenology to the conclusion that phenomenal consciousness does not depend upon sensory processing in the brain? The short answer is that P2 is the result of an *inference to the best explanation*. Namely, the best explanation for why CP does not depend upon SP is that sensory processing is not a necessary condition for phenomenal consciousness.

⁵ Brown & Mandik (2012) also make the point that consistency demands that first-order cognitive states be made phenomenally conscious when discussing HOT theory in general.

The defender of P2 can also point out that it is commonplace in the philosophy of mind to draw metaphysical conclusions about the nature of consciousness based upon phenomenological premises. Two examples of such arguments are the ‘Transparency Argument for First-Order Representationalism’ (Tye, 2002) and the ‘Peripheral Inner Awareness Argument for Self-Representationalism’ (Kriegel, 2009). Both of these arguments assume that the phenomenology of conscious experience can inform us about the underlying metaphysical nature of experience. With respect to the ‘Transparency’ argument, for instance, the idea is that first-order representationalism (a metaphysical thesis) can be derived from the fact that the only introspectively discernible aspect of experience is its world-directed representational content. P2 of the cognitive phenomenology argument for disembodied AI consciousness employs similar reasoning. Specifically, the premise relies on the following principle (call it ‘PHEN’):

PHEN: Phenomenological independence implies metaphysical independence.

According to PHEN, it follows from the fact that X is phenomenologically independent of Y to the conclusion that X can be instantiated in the absence of the metaphysical underpinning of Y. Applied to cognitive phenomenology, the idea is that if CP is independent of SP, then CP can be instantiated in the absence of sensory processing, meaning that sensory processing is not a precondition for phenomenal consciousness.

One might object to PHEN on the grounds that there are obvious counterexamples to the principle. For example, <pain> is phenomenologically independent of <C-fibers firing> given that the latter is not phenomenological at all. It does not follow from this fact, however, that <pain> is metaphysically independent of <C-fibers firing>, or that the sensation of pain is not equivalent to the firing of C-fibers.⁶ <Pain> may be metaphysically dependent upon <C-fibers firing> (i.e., the *thesis of multiple realizability* may be false), despite the fact that the former is phenomenologically independent of the latter.

This objection to PHEN rests on a (understandable) misunderstanding of what is meant by ‘phenomenological dependence/independence’. The relationship of phenomenological dependence/independence is by stipulation a relationship that holds between two *phenomenal* properties, signifying that the relata must both be experientially manifest. Since the firing of C-fibers is not experientially manifest, it makes no sense to say that <pain> is phenomenologically independent of <C-fibers firing>. It does make sense, however, to say that agential experience is phenomenologically independent of moral experience, or that conative experience is phenomenologically independent of cognitive experience, or that cognitive experience is phenomenologically independent of sensory experience. For all of the relata in these cases are experientially manifest properties.

Of course, PHEN remains controversial even with this clarification in place. It may be the case, for example, that while sensory processing (e.g., the presence of what Koch calls ‘the sensory hot zone’ (Koch, Massimini, Boly & Tononi, 2016)) is a developmental precondition for consciousness, it is nevertheless true that as we get older, we develop a kind of sui generis cognitive phenomenology that can be instantiated in the absence of sensory phenomenology.

⁶ I thank Bill Lycan for bringing this objection to my attention.

There is, on the face of it, nothing conceptually problematic about this objection to P2. The burden of proof is squarely on the objector though to explain how it is possible that (a) phenomenal consciousness metaphysically depends upon sensory processing, but (b) cognitive experience does not phenomenologically depend upon sensory experience. I am not aware of any explanation to this effect that currently exists in the literature. P2 is highly plausible until such an explanation is given.

Before moving on to P3 of the cognitive phenomenology argument, it is worth drawing attention to the fact that weak primitivism about CP clearly does not indicate that sensory processing is not necessary for consciousness. This is precisely because weak primitivism states that cognitive phenomenology cannot be instantiated in the absence of sensory phenomenology (despite being irreducible to SP).

5. Premise 3: From the Non-Necessity of Sensory Processing to Disembodied AI Consciousness

Assume now that both P1 and P2 are true, and that consequently, phenomenal consciousness does not require sensory processing. How does this conclusion bear on the philosophy of artificial intelligence? In particular, what is the reasoning behind P3? It is vital to accentuate that P3 operates under the controversial assumption that AI consciousness is in principle possible. I do not make any arguments for or against the possibility of AI consciousness here. Rather, I simply assume that AI consciousness is possible and proceed to address the following question (call it ‘the embodiment question’ (EQ)):

EQ: Would AI consciousness, should it exist, require embodiment?

EQ is a pressing question in the philosophy of artificial intelligence. We want to know whether or not disembodied AI systems (such as neurotrophic supercomputers or sophisticated personal assistants – perhaps future versions of Apple’s ‘Siri’ or Google’s ‘Alexa’) are capable of what-it’s-likeness.⁷ It is likely, after all, that such systems will be able to pass the Turing test in the not too distant future. According to a survey conducted by Bostrom and Muller (2016), “experts estimate the probability of achieving ‘high level machine intelligence’ to go over 50% by 2040-2050, over 90% by 2075” (Muller, 2016: 4).

One relevant worry here is that we will *inevitably attribute* consciousness to future disembodied AIs. Perhaps it is psychologically infeasible to deny phenomenal experience to an AI system that is verbally indistinguishable from a normal adult interlocutor. If this is the case, and disembodied AIs are *incapable* of experience, then there is a very real concern that humanity will be psychologically duped into thinking that unfeeling computer programs are phenomenally conscious. I call this the *duped-dystopia scenario*. An answer to EQ must be provided in order to ensure that the duped-dystopia scenario is avoided. P3 states that the answer to EQ is no, assuming that phenomenal consciousness does not require sensory processing. The basic proposition motivating P3 is that *embodiment is necessary for consciousness only if sensory processing is necessary for consciousness*. To resist this proposition, the objector to P3 must

⁷ It is not clear whether deep learning systems count as ‘sensory’ in the same way that a brain counts as sensory.

somehow demonstrate that consciousness requires embodiment even if consciousness does not necessitate sensory processing. It is *prima facie* difficult to discern how this feat could be accomplished.

The first point to note here is that ‘embodiment’ is often understood (at least partly) in terms of the ability to exercise sensorimotor capacities. Take the *embodied cognition approach* in cognitive science, for example. As indicated by its name, the embodied cognition approach maintains that the instantiation of cognitive processes requires embodiment. The approach thereby rejects the traditional computational view of the mind, according to which cognition is essentially a matter of internal information processing. Crucially, proponents of the embodied cognition approach typically reference sensorimotor capacities when defining the concept of embodiment. Monica Cowart, for example, asserts that within the context of the embodied cognition debate, “the basic notion of embodiment is broadly understood as the unique way an organism’s sensorimotor capacities enable it to successfully interact with its environmental niche” (Cowart, 2004). In a similar vein, Chris Letheby defines the embodied cognition approach as the thesis which “holds that conceptual and linguistic thought is grounded in the brain’s perceptual and sensorimotor systems” (Letheby, 2012: 403). If embodiment by definition requires the possession of sensorimotor capacities (and hence sensory processing), then P3 is necessarily true. For it is impossible that embodiment is a precondition for consciousness if embodiment requires sensory processing, and it is stipulated that sensory processing is *not* a precondition for consciousness.

I will assume for the sake of argument that embodiment does not by definition require the possession of sensorimotor capacities. Even granting this assumption, the burden of proof is still on the objector to P3 to explain how embodiment could be necessary for phenomenal consciousness given that consciousness does not depend upon sensory processing. There are different ways the objector might go about providing such an explanation. From my perspective, the only feasible way to object to P3 is to make the following threefold contention:

Contention 1: *Being alive* is a necessary condition for being phenomenally conscious;

Contention 2: Embodiment is a necessary condition for *being alive*;

Contention 3: Being phenomenally conscious does not require the possession of sensorimotor capacities.

Successfully making this threefold contention is synonymous with illustrating that consciousness requires embodiment but not sensory processing, and is thus synonymous with illustrating the falsity of P3. There is, of course, the question as to whether these three contentions are *mutually consistent*. The position in the philosophy of consciousness known as *autopoiesis enactivism* is committed to both Contention 1 and Contention 2. It is unclear, however, whether autopoiesis enactivism is compatible with Contention 3. I do not consider any arguments for or against autopoietic enactivism here. Rather, I assume that autopoietic enactivism is plausible and advance to ask whether the view can be weaponized as an objection to P3. Making sense of this question requires an understanding of enactivism, and in particular, an understanding of the distinction between *autopoiesis enactivism* and *sensorimotor enactivism*.

Enactive theories of consciousness understand consciousness to be interactive in nature as opposed to passive, meaning that phenomenal experience according to the enactivist is

something that we do as opposed to something that transpires within us. As Jan Degenaar and J. Kevin O' Regan observe, "Enactive approaches to consciousness emphasize patterns of interaction, and active engagement, instead of any alleged internal representations of the environment" (Degenaar & Regan, 2017: 393).

Enactivism represents an embodied theory of consciousness due to its emphasis on bodily interactions with the external world.⁸The pressing question for any enactive theory is: exactly which interactive capacities are necessary and sufficient for consciousness? This inquiry raises the crucial distinction between *autopoietic* enactivism and *sensorimotor* enactivism. The latter version of the view maintains that the relevant interactive capacities are sensorimotor capacities. More to the point, sensorimotor enactivism says that the possession of sensorimotor capacities is both necessary and sufficient for consciousness (O'Regan and Noë, 2001; O'Regan, 2011; Degenaar and O'Regan, 2015). By contrast, autopoietic enactivism de-emphasizes the role of sensorimotor capacities, focusing instead on a perceived inextricable link between consciousness and interactive capacities associated with being alive. 'Life' in the context of this theory is defined in terms of 'autopoiesis'. As Degenaar and Regan clarify, "In the enactive literature, life is typically understood as autopoiesis (Maturana & Varela, 1980). Living organisms are spontaneously active, self-maintaining and self-producing systems, and it has been proposed that we should understand our phenomenology in light of this self-creating or 'autopoietic' organization" (Degenaar & Regan, 2015: 396). Degenaar and Regan further elucidate that 'autopoiesis' necessarily involves embodied interaction with the environment: "An autopoietic system needs to metabolize and continuously adapt to its environment in order to continue its activity. By its very definition, autopoiesis cannot withdraw itself from the interactive context" (Degenaar & Regan, 2015: 397-398).

Autopoietic enactivism, then, says that life is necessary for consciousness, where life is taken to be a fundamentally interactive process that mandates embodiment. It is for this reason that autopoietic enactivism is committed to Contention 1 and Contention 2 above. The essential question for the objector to P3, as indicated, is whether autopoietic enactivism is compatible with Contention 3, according to which the possession of sensorimotor capacities is not necessary for consciousness. There is reason to doubt that the two are compatible. Autopoietic enactivism is, after all, motivated by the conviction that "perceptual consciousness involves *additional* constraints over and above perceptual capacity" (Degenaar & Regan, 2015: 397-398). In other words, while autopoietic enactivists deny that the possession of sensorimotor capacities is a *sufficient condition* for consciousness, they tend to agree with sensorimotor enactivists that possessing said capacities is a *necessary condition* for consciousness. Is it possible for the autopoietic enactivist to uphold Contention 3, even though the majority of autopoietic enactivists explicitly accept the proposition that consciousness requires sensorimotor skills? I would argue that this answer to this question is no. It seems to me that autopoietic enactivism and Contention 3 are mutually inconsistent because autopoietic processes arguably depend upon sensorimotor skills.

⁸ The concept of 'enaction' derives from the book *The Embodied Mind* (Varela, Thompson and Rosch, 1991), in which the authors attempt to revolutionize cognitive science by connecting cognitive science with Merleau-Ponty's (1945) phenomenological work on embodiment.

To summarize: P3 states that AI consciousness (assuming that it is possible) does not require embodiment if phenomenal consciousness does not require sensory processing. To object to P3, one must explain how embodiment could be necessary for phenomenal consciousness, given that consciousness does not depend upon sensory processing. I first pointed out that it is impossible for the objector to achieve this feat if embodiment is defined (at least partly) in terms of the ability to exercise sensorimotor skills. Then, I considered and eventually rejected the idea that autopoietic enactivism can be levied as an objection to P3.

6. Conclusion

P1 of the CP argument for disembodied AI consciousness holds that strong primitivism about CP is true. I have not attempted to establish the veracity of strong primitivism here. Instead, I have merely contended that strong primitivism is credible given certain theoretical assumptions; in particular, the assumption that pure consciousness is possible and/or the assumption that Rosenthal's HOT theory is true. P2 of the argument goes on to state that sensory processing is not a necessary condition for phenomenal consciousness if strong primitivism is true. I argued that P2 is strongly supported by what I call PHEN, but acknowledged that PHEN can be reasonably resisted. Finally, I illustrated that while it is possible to object to P3, it is fair to say that P3 is the least controversial premise of the argument. The conclusion that AI consciousness does not necessitate embodiment is by no means trivial. Such a conclusion widens the scope of possible types of conscious AIs, and arguably even suggests that there are two fundamentally different kinds of consciousness: sensory- based consciousness and disembodied consciousness.

References

- Block, Ned (1995) “On a confusion about a function of consciousness”, *Brain and Behavioral Sciences*, 18 (2): 227-247.
- Block, Ned (2011) “Perceptual consciousness overflows cognitive access”, *Trends in Cognitive Sciences*, 15 (12): 567-575.
- Block, Ned (2011) “The higher order approach to consciousness is defunct”, *Analysis*, 71 (3): 419-431.
- Brentano, Franz (1874) *Psychology From an Empirical Standpoint*, New York: Routledge.
- Brown, Richard & Mandik, Pete (2012) “On Whether the Higher-Order Thought Theory of Consciousness Entails Cognitive Phenomenology, or: What is it Like to Think that One Thinks that P?”, *Philosophical Topics*, 40 (2): 1-12.
- Carruthers, Peter (2005) *Consciousness: Essays From a Higher-Order Perspective*, Oxford: Oxford University Press.
- Carruthers, Peter (2008) “Higher-order theories of consciousness” In *Stanford Encyclopedia of Philosophy*.
- Carruthers, Peter & Veillet, Bénédicte (2011) “The Case Against Cognitive Phenomenology” In *Cognitive phenomenology* (Tim Bayne & Michelle Montague, eds.), Oxford: Oxford University Press, pp. 35.
- Chudnoff, Elijah (2015) “Phenomenal Contrast Arguments for Cognitive Phenomenology”, *Philosophy and Phenomenological Research*, 90 (2): 82-104.
- Coleman, Sam (2015) “Quotational higher-order thought theory”, *Philosophical Studies*, 172 (10): 2705-2733.
- Cowart, Monica (2004) “Embodied cognition” In *Internet Encyclopedia of Philosophy*.
- Degenaar, Jan & O’Regan, J. Kevin (2017) “Sensorimotor Theory and Enactivism”, *Topoi*, 36 (3): 393-407.
- Gennaro, Rocco J. (2018) “Higher-Order Theories of Consciousness” In *Internet Encyclopedia of Philosophy*.
- Horgan, Terence & Tienson, John (2002) “The Intentionality of Phenomenology and the Phenomenology of Intentionality” In *Philosophy of Mind: Classical and Contemporary Readings* (David J. Chalmers, ed.), Oxford: Oxford University Press, pp. 520–533.
- Katz, Steven T. (1978) “Language, epistemology, and mysticism” In *Mysticism and Philosophical Analysis*, Oxford: Oxford University Press, pp. 22-74.

- Koch, Christof & Massimini, Marcello, Boly, Melanie & Tononi, Giulio (2016) “Neural correlates of consciousness: Progress and problems”, *Nature Reviews Neuroscience*, 17: 307-321.
- Kriegel, Uriah (2009) “Self-representationalism and phenomenology”, *Philosophical Studies*, 143 (3): 357-381.
- Letheby, Christopher (2012) “In defence of embodied cognition: a reply to Fred Adams”, *Phenomenology and the Cognitive Sciences*, 11 (3): 403-414.
- Levine, Joseph (2011) “On the Phenomenology of Thought” In *Cognitive Phenomenology* (Tim Bayne and Michelle Montague. eds.), Oxford: Oxford University Press, pp. 103.
- Lycan, William G. (1996) *Consciousness and Experience*, Cambridge, MA: MIT Press.
- Mendelovici, Angela (2013) “Intentionalism about Moods”, *Thought: A Journal of Philosophy*, 2 (1): 126-136.
- Merleau-Ponty, Maurice (1962) *Phenomenology of Perception*, New York: Routledge.
- Rosenthal, David (2005) *Consciousness and Mind*, Oxford: Oxford University Press.
- Rosenthal, David M. (1993) “Higher-order thoughts and the appendage theory of consciousness”, *Philosophical Psychology*, 6 (2): 155-66.
- Rosenthal, David M. (1997) “Perceptual and cognitive models of consciousness”, *Journal of the American Psychoanalytic Association*, 45.
- Rosenthal, David M. (1986) “Two concepts of consciousness”, *Philosophical Studies*, 49 (May): 329-59.
- Rosenthal, David (2004) “Varieties of higher-order theory” In *Higher-Order Theories of Consciousness: An Anthology* (Rocco J. Gennaro, ed.), Amestradam: John Benjamins.
- Smith, David (2011) “Nibbanic (or Pure) Consciousness and Beyond”, *Philosophia*, 39 (3): 475-491.
- Tye, Michael (2002) “Representationalism and the transparency of experience”, *Noûs*, 36 (1): 137-51.
- Varela, Francisco, Thompson, Evan & Rosch, Eleanor (1991) *The Embodied Mind: Cognitive Science and Human Experience*, Cambridge, MA: MIT Press.

