# Predictive processing and extended consciousness: why the machinery of consciousness is (probably) still in the head and the DEUTS argument won't let it leak outside

**Abstract:**

Recently, Kirchhoff and Kiverstein have argued that the extended consciousness thesis, namely the claim that the material vehicles of consciousness extends beyond our heads, is entirely compatible with, and mandated by, the predictive processing framework. To do so, they rely on a potent argument in favor of the extended consciousness thesis, namely the Dynamical Entanglement and Unique Temporal Signature (DEUTS) argument. Here, we will critically examine Kirchhoff and Kiverstein's endeavor, arguing for the following three claims. First, we will claim that Kirchhoff and Kiverstein's emphasis on culture and cultural practices does not help them substantiate the extended consciousness thesis. Secondly, we will argue that the way in which Kirchhoff and Kiverstein formalize the boundaries of a subject's conscious mind is inadequate, as it yields conclusions running counter some of their assumptions. Lastly, we will argue that the DEUTS argument does not establish the extended consciousness thesis, as it licenses a "consciousness bloat" objection which is exactly analogous to the "cognitive bloat" objection to the extended mind thesis. We will thus conclude that Kirchhoff and Kiverstein's proposed marriage between the extended consciousness thesis and predictive processing fails, and that, contrary to a popular opinion, DEUTS is not a strong argument in favor of the extended consciousness thesis.

**Keywords:** Predictive Processing, Free-Energy Principle, Markov Blankets, Extended Consciousness, Sensorimotor Enactivism.

## 1 - Introduction

The extended consciousness thesis, or *consciousness vehicle externalism* (CVE) claims that the *material vehicles* of a subject's consciousness[1] at least sometimes include the subject's active body and/or some appropriate environmental prop (Hurley 2010; Vold 215).

Sensorimotor enactivists endorse CVE. They claim that phenomenally conscious perception is *achieved* by an embodied agent by interacting sensomotorically with the environment; for instance, by saccading over a visual target so as to explore its visual profile. They argue that the phenomenal qualities of perception are determined by *sensorimotor contingencies*: law-like relations holding between bodily movements and changes in sensory

[1] We will use "consciousness" to refer to *phenomenal* consciousness throughout the paper.

stimulation (O'Regan and Noë 2001a; 2001b; Hurley and Noë 2003; O'Regan 2011). On their view, perceivers possess a tacit body of knowledge concerning such relations, called sensorimotor mastery, that they exert so as to *enact* their perceptual experiences; that is, perceptually explore the environment so as to reveal it phenomenally. The vehicles of perceptual phenomenology thus include, alongside patterns of neuronal activation, bodily acts of perceptual exploration guided by the perceiver sensorimotor mastery (Noë 2004, 2009; Kiverstein and Farina 2012; Pepper 2014).

This formulation of CVE has recently been put under significant pressure by the neurocomputational framework of *predictive processing* (PP). This is because PP allows to operationalize sensorimotor contingencies in purely neural terms, as expectations concerning the incoming inputs encoded in a generative model (e.g. Pezzulo *et al* 2017; Baltieri and Buckley 2019). The perceiver mastery of sensorimotor contingencies is thus "pushed inside the brain", and ends up regulating only neuronal message-passing, leaving no reason to endorse CVE (Clark 2012; Seth 2014).[2]

Recently, Kirchhoff and Kiverstein (2019a; 2019b; 2020) set off to revise this dialectical situation, showing that a properly understood PP *mandates* CVE (Kirchhoff 2018; Kirchhoff and Kiverstein 2019a: 57-59; 83-86). To do so, they carefully analyze the conceptual apparatus of PP, and use it to formulate the most potent argument in favor of CVE in the sensorimotor enactivists' arsenal; namely the Dynamical Entanglement and Unique Temporal Signature (DEUTS, see § 3 below) argument.

Here, we diagnose three problems in Kirchhoff and Kiverstein's proposal. First, we argue that their appeal to cultural practices does little to support CVE. Secondly, we argue that their usage of Markov Blankets (more on which in §2) to formalize the boundaries of a subject's

[2] Importantly, when issues regarding CVE are left aside, PP and sensorimotor enactivism have a far less hostile relation both conceptually (e.g. Vázquez 2020) and empirically (e.g. Laflaquiere 2017; Leinweber *et al*. 2017).

conscious mind yields results that clash with some of their theoretical commitments. Lastly, we argue that the DEUTS argument generates a "consciousness bloat" problem analogous to the "cognitive bloat" problem affecting the extended mind thesis (see Sprevak 2009; Rowlands 2010). We will thus conclude that Kirchhoff and Kiverstein's endeavor fails to secure a happy marriage between CVE and PP, and that, contrary to a popular opinion (Clark 2009; 2013; Kirchhoff and Kiverstein 2019a), the DEUTS argument is not a *strong* argument in favor of CVE.

Here's our plan. In the next two sections, we will sketch PP and Kirchhoff and Kivertein's position respectively. In section four, we will turn from exposition to criticism, articulating our three claims. A brief concluding paragraph follows.

## 2 - Predictive processing and the free-energy principle: a quick introduction

PP is a neurocomputational framework providing a process theory for Friston's free-energy principle (Friston and Stephan 2007; Hohwy 2020). We succinctly introduce them here.[3] Readers already familiar with these frameworks might wish to skip ahead.

The *free-energy principle* states that biological self-organization is centered around the avoidance of *surprisal*: an information-theoretic measure of the unexpectedness of sensory states, given a model (implicitly realized by the organism's embodiment) of the sensory states compatible with the organism's prolonged existence, which the organism should expect to occupy (Friston 2012a; 2013). Organisms, however, cannot track surprisal directly. They can only track its upper bound, which is (variational) free-energy (Friston 2009). Free-energy is an upper bound on surprisal because it can be understood as surprisal plus a second, *always positive*, quantity, which is the Kullback-Leibler Divergence ($D_{KL}$): a measure of how much

[3] For more introductory material, see (Hohwy 2013; Clark 2016; Tani 2016; Wiese and Metzinger 2017).

the system's "guesses" about the causes of its states are aligned with reality. Since surprisal is also the complement of model evidence (Friston 2019: 177), minimizing it amounts to producing the evidence in favor of one's model of one's prolonged existence. Surprisal-minimizing systems are thus self-evidencing systems (Hohwy 2016): systems striving to bring about the evidence favoring the hypothesis that they exist, thereby prolonging their existence.

In the context of the free-energy principle, such models are understood as the states enclosed by a Markov Blanket (Hohwy 2017; Friston et al. 2020). Markov Blankets are *formal* boundaries that separate a biological system from its niche in a statistical sense, while allowing the two to causally interact (Friston 2013; Friston et al. 2020).[4] The interaction is allowed by the internal partition of the Markov Blanket: active states allow the organism to influence its niche, whereas sensory states allow the niche to influence the organism. In this way, the Markov Blanket allows a two-way causal interaction between organism and niche, enabling the coupling between the two (Hohwy 2017; Kirchhoff and Kiverstein 2019a: 65-67).

Importantly, according to the free-energy principle, Markov Blankets are multiple and nested, and can be found at each and every level of organization (Kirchhoff et al. 2018). A biological system such as a wolf, for instance, can be decomposed in biological sub-systems (e.g. cells) and can partake in larger systems (e.g. a pack), each busy minimizing free-energy (and thus, equipped with Markov Blankets).

The free-energy principle can be related to PP by noticing that free-energy can be equated

[4] There is a vast philosophical literature on Markov Blankets, detailing how this conception of Markov Blankets is removed from the one relevant in machine learning (e.g. Bruinenerg et al. 2020). There is also a current controversy on the ontological status of Markov Blankets, concerning whether they are objective features of biological systems or just modelling tools (see Menary and Gillet 2020; Andrews 2021). Here, we will assume for the sake of discussion that the realistic reading is correct, as it seems the one espoused by Kirchhoff and Kiverstein (2019a; 2019b).

(under some assumptions, see Buckely *et al*. 2017) with prediction error: a well-known *neural* signal posited by predictive coding accounts of neural functioning (e.g. Rao and Ballard 1999, Friston 2005). In the context of PP, prediction error names the *mismatch* between endogenously generated sensory states, which are the ones the agent expects or predicts, and actually received ones. On the view of the mind PP proposes, the primary function of brains is to minimize prediction errors, and cognition unfolds as a consequence of error minimization.

Prediction error can be minimized in two ways (Hohwy 2013; Clark 2016). One is by revising the expectations, fitting them to the received sensory signals. This is perceptual recognition, which, in free-energy terms, corresponds to decreasing the $D_{KL}$. Another way to minimize prediction error is by changing the received sensory states through movement, so as to bring about the expected ones. This is active inference, which, in free-energy terms, corresponds to a direct minimization of surprisal.

Importantly, prediction errors are always weighted based on their expected precision: roughly, an estimate of the signal to noise ratio of the sensory signals giving rise to prediction errors. In this way, sensory signals generated by highly noisy sensory states will be dampended, thereby allowing only informative prediction errors to dominate neural processing (Feldman and Friston 2010).

Active inference puts PP in direct contact with sensorimotor enactivism. This is because active inference is performed by predicting the sensory consequences of one's movement, and then canceling out the error they generate through movement (Friston 2011; Adams *et al*. 2013). Hence, on the account of action predictive processing offers, to perform a movement *m* an agent must first predict the sensory outcomes *m*, which are the sensory states the agent would encounter, were *m* performed. So, on the account of action PP offers, the agent must know *how* actions systematically impact the incoming sensory stream; that is, the agent must *know* the relevant sensorimotor contingencies, and exert their mastery by predicting a "desirable" stream of sensory inputs.[5]

Notice how this makes an agent's mastery of sensorimotor contingencies a purely *neural* affair (Seth 2014), pushing us towards an indirect and (vehicle) internalist view of perception (Wiese 2018). If the perceiver's mastery of sensorimotor contingencies is neurally realized, and sensorimotor contingencies only mediate the message passing in a neurally realized generative model, then perception naturally appears an *internal* affair.

Notice, lastly, that PP is *not* a theory of consciousness (phenomenal or otherwise), even if it seems able to account for at least some structural aspects of consciousness and might be relevant for the empirical research on consciousness (see Seth and Hohwy 2020). Kirchhoff and Kiverstein, however, seem to take PP as a theory of consciousness. For instance, they write:

> "Predictive processing tells us what the parts of the system must be doing such that when these parts are organised in the right way, they constitute consciousness. The parts of the system will include, for instance, components that perform predictions, error calculation, precision estimation, and so on." (Kirchhoff and Kiverstein 2019a: 104).

We concede the point for the sake of argument. Crucially, however, Kirchhoff and Kiverstein hold that only *temporally thick* generative models qualify as consciousness supporting (Kirchhoff and Kiverstein 2019a: 106-108, see also Hobson and Friston 2014). Here, we will assess the temporal thickness of generative models using the following heuristic: the more the model allows a system to make larger loops in the space of all its possible states, the temporally thicker it is (Friston 2018: 5-6). For instance, a model that allows a subject to

[5] Notice that in more traditional theories of motor control such a role is played by *forward models* (also called motor emulators), which have already been used to operationalize sensorimotor contingencies (e.g. Maye and Engel 2013).

celebrate her birthday on an annual basis is considerably thicker than the model of an agent that loops through its state space every few seconds. This way of proceeding seems to fit quite nicely with Kirchhoff and Kiverstein's (2019a: 55-57) understanding of models, according to which models are first and foremost *entire organisms*.[6]

But what are temporally thick models *made of*? What is the *machinery* constituting them? PP suggests that this machinery is squarely located within brains. Kirchhoff and Kiverstein disagree. Let us see why.

## 3 - The DEUTS argument, twenty(ish) years later

DEUTS is often considered the strongest argument for CVE (Clark 2009; 2013; Kirchhoff and Kiverstein 2019a). Its origins trace back to (Hurley 1998: Ch.8). There, she makes a two-stepped argument. Here, we briefly present each step in its original variant, followed by Kirchhoff and Kiverstein's PP rendition of it.

### 3.1.1 - The first step: Dynamical Entanglement

The first step claims that cognitive processing sometimes weaves agent and environment in a single system. This claim stems from dynamical approaches to cognition, according to which cognitive processing is not "sandwiched" between perception and action (Hurley 2001), but rather *constituted* by cyclical sensorimotor interactions. These interactions are best

[6] There are other ways to determine the temporal thickness of a model. One is simply that of observing its hierarchical structure, in which hierarchically higher layers generate predictions over longer timescales, and are thus temporally thicker than hierarchically lower ones. But this way of proceeding seems to apply only to hierarchically structured internal (brain-like) models (e.g. Tani 2016: Ch. 9-10). Hence, it would beg the question against Kirchhoff and Kiverstein (2019a: 55) understanding of models. Another way to capture the temporal thickness of a model may be in terms of expected free energy (Corcoran et al. 2020): roughly, the free-energy expected in the future, after active inference has been performed. However, expected free-energy might not accurately model the free-energy expected in the future, and hence it might prove inadequate to capture the temporal thickness of a model (Millidge et al. 2021). We rely on Friston's (2018) heuristic to avoid these problems.

explained using the formal tools of dynamical systems theory (e.g. Hurley 1998; Chemero 2009) which allow to quantitatively model and predict them. This explanatory methodology, however, often forces one to model agent and environment as a single (non-decomposable) coupled system, whose behavior can be accounted for only by using its *order parameters*[7], and whose dynamics accounts for the production of cognitive outputs (Lamb and Chemero 2018). Hence, in those cases, agent and environment form a single cognitive system, which vindicates a form of vehicle externalism about *cognition* (Palermos 2014; Kiverstein 2018). How can this dynamical image of (extended) cognition be related to PP?

### 3.1.2 - Dynamical Entanglement and Predictive Processing.

Kirchhoff and Kiverstein commence by noticing that prediction error minimization is a tool for surprisal avoidance. But surprisal can be avoided only through active inference[8]; that is, embodied action. Thus within PP, real, embodied action is central to cognition (Kirchhoff and Kiverstein 2019a: 57-59).

Secondly, Kirchhoff and Kiverstein notice that albeit Markov Blankets statistically separate agent and environment, they also *enable* the coupling of the two (Kirchhoff and Kiverstein 2019a: 65-67; see also Fabry 2017). This is due to the interplay of the active and sensory states that jointly constitute the blanket. Recall: active states *influence* sensory and external states, and are *influenced by* internal states. Conversely, sensory states *influence* active and internal states, and are influenced by external states.[9] Thus, together, active and sensory states enable internal and external states to interlock in a *two way* interaction, which

[7] An order parameter (or collective variable) is a variable describing the behavior of the components of the entire system. See (Kelso 1995, Ch 1 and 2) for a useful introduction.
[8] Notice that here we are dealing with surprisal. *Free-energy* can be minimized in other ways too, as indicated above.
[9] Notice, importantly, that sensory and active states influence each other, and are thus coupled. This, on Kirchhoff and Kiverstein's (2019a: 69) view, allows Markov Blankets to capture the idea of sensorimotor contingencies.

is a form of coupling.

In third place, relying on Clark's (2017) metamorphosis argument, Kirchhoff and Kiverstein stress that Markov Blankets are not just multiple and nested, but also *malleable and plastic*. Clark invites us to consider metamorphic insects, and the boundary (i.e. the Markov Blanket) separating them from the environment. As the insect undergoes the metamorphic process, it re-negotiates that boundary, shifting the set of states that separates it from the environment (trivially, the silk a cocoon is made of is not the exoskeleton of the caterpillar). Hence, Markov Blankets are not fixed: self-evidencing models such as metamorphic insects may undergo dramatic changes, thereby shifting the Blanket separating them from their niche. To this, Kirchhoff and Kiverstein add that, at least sometimes, these shifts can "push inside the Blanket" some environmental prop essential to the self-evidencing of an agent. This is the case of a spider and its web, or of a mildly amnesic patient and the notebook he uses to compensate for his amnesia (Kirchhoff and Kiverstein 2019a: 73-76; 2019b).

Consider this point in the light of the coupling Markov Blankets enable. If an agent can be dynamically entangled with an external resource, they form a single coupled dynamical system. And if it avoids surprisal (e.g. Bruineberg 2018b) it will be a free-energy/prediction error minimizing system in its own right, with its own Markov Blanket. In such a case, the coupled system will be identified through a "wider" Markov Blanket, encompassing "smaller" coupled Markov Blankets. The "wider" Markov Blanket will act as an *order parameter* on the "smaller" Blankets it contains, constraining their degrees of freedom and coalescing them in a single, possibly extended, system (Kirchhoff and Kiverstein 2019a: 80-81). In Kirchhoff and Kiverstein's view, such a Blanket identifies the relevant cognitive machinery, namely, the relevant self-evidencing model engaged in prediction error/free-energy minimization. They hold that, by default, such a model encompasses the entire organism (Kircchoff and Kiverstein 2019a: 55-57), but, as the metamorphosis argument purportedly shows, it can extend to incorporate the environmental props the system is coupled to, allowing the creation "on the spot" of extended free-energy/prediction error minimizing systems. In this way, Kirchhoff and Kiverstein resort to Markov Blankets to formalize the boundaries of (extended) cognitive systems.

Here, we summarized Kirchhoff and Kiverstein's PP inspired rendition of the first step of the DEUTS argument – the "Dynamical Entanglement" bit. Yet, thus far, the argument seems only an argument for vehicle externalism about *cognition*. Where does consciousness come into play? To answer this question, we need to look into the second step of the DEUTS argument – the "Unique Temporal Signature" bit

### 3.2.1 - The second step: Unique temporal Signature

The idea behind the second step of the argument is that a careful consideration of the temporal development of our experiences (their "unique temporal signature") shows that the dynamical avenue to vehicle externalism about cognition *entails* CVE.

To claim so, Hurley (1998: Ch. 8) starts by granting the familiar vehicle internalist intuition that *if* the machinery of consciousness is purely neural, *then* keeping a subject's neural state constant will keep the subject's phenomenology constant, *whatever the environment*. Hence, if the machinery of consciousness is purely neural, two subjects can be neural and phenomenal duplicates *without* being environmental duplicates.

To attack it, Hurley (1998: Ch.8) argues that two *dynamically entangled* subjects cannot be neural and phenomenal duplicates without being also *environmental* duplicates. To see why, consider a simplified rendering of one of Hurley's (1998: 303-314) thought experiments. On earth, subject S is in an entirely white room, containing only S and a black ball on S's right. On twin earth, S's physical duplicate TS is in the exact same situation. Since S and TS are physical duplicates, it seems correct to say S and TS are experiencing the same thing: what it feels like to be in a white room with a black ball at one's right. Here, S and TS are phenomenal, neural and environmental duplicates. But (and this is the vehicle internalist intuition) S and TS *could* be phenomenal duplicates *just* by being neural duplicates: we could switch the place of TS's ball from right to left and insert in TS's eyes left-to-right inverting lenses, so as to keep the visual input TS receives constant (Hurley 1998: 304). In this case, it seems correct to say that TS and S will be in the same neural state, undergoing the same experience.

Hurley (1998: 327) suggests this is possible only because S and TS are *passive perceivers*, and thus they are not coupled with the environment in any meaningful way. Were S and TS dynamically entangled with their (different) environments, they *would* cease to be phenomenal duplicates. Suppose, for instance, that S and TS try to touch the ball when they're not environmental duplicates. They will both move their right arms towards the ball (which they both *see* at their right), causing their neural and phenomenal states to diverge. For S will touch the ball, whereas TS will not. So only S'n neural state will be modified by the reafferent signals. As a result, only S will experience what it is like to touch a ball. In this case, phenomenal duplication *fails*.

To duplicate the experience of the dynamical entangled S to be duplicated in TS, one needs to make S and TS *environmental duplicates*; that is, one has to remove the lenses from TS's eyes and displace TS's ball in its original position. Only in this case S's temporally extended experience can be duplicated in TS. So, in order for the phenomenology of a dynamically entangled subject to be duplicated, it is not sufficient that the subject and their twin are neural duplicates. They also *need* to be environmental duplicates.[10]

But what needs to be the case in order for a phenomenal state to occur is the *vehicle* of that state (Hurely 1998: 330-331). And, in the example just considered, what needs to be the case for the relevant phenomenal states to occur includes at least some environmental factors. Hence, if a subject is dynamically entangled, environmental factors are vehicles of its consciousness (Hurley 1998: 330-335). Reflections on the unique temporal development of experiences thus show that dynamical entanglement *entails* CVE. Importantly, such a procedure to identify the external vehicle of consciousness is supposed to be *discriminating* (Hurley 1998; 330-331); that is, it is supposed to be able to tell apart genuine external vehicles of a subject's consciousness from external factors merely impacting it.

### 3.2.2 - Unique Temporal Signature and Predictive Processing

Kirchhoff and Kiverstein (2019a: 36; 112-115; 2020: 5-9) build essentially on the same point. Yet they would emphasize the role of the *cultural* environment in the constitution of one's experience (Kirchhoff and Kiverstein 2019a: Ch. 5 and 6; 2020). This is because they adhere to a "third wave" form of vehicle externalism, asserting that the cognitive (and phenomenal) machinery has no fixed properties. Rather, its properties are constantly transformed by the cultural practices[11] the subject masters, and these properties have to be constantly negotiated by engaging with the surrounding cultural niche (Kirchhoff and

[10] Importantly, the environmental duplication need not be total:only the environmental factors with which S is dynamically entangled *need* to be duplicated in TS. Non-phenomenology-shaping aspects of the environment might still systematically differ.

[11] To our knowledge, Kirchhoff and Kiverstein do not provide an explicit definition of cultural practices. However, it seems that they interpret cultural practices as *relatively stable, socially regimented, ways of interaction*, which induce a relatively stable sensory flow to which an agent can attune its expectations (cfr. Kirchhoff and Kiverstein 2020). This way of understanding cultural practices might derive from Kiverstein's past work on the skilled intentionality framework, which explicitly understands practices as "stable ways of doing things" (Rietveld and Kiverstein 2014; see also Rietveld *et al*. 2018 for a presentation of the framework). Many thanks to **a colleague** for having pointed this out.

Kiverstein 2019a: Ch. 5; 2020). Subjects are thus enculturated, in the sense that their participation in cultural practices and their attunement to their cultural niches transforms them and alters their properties.

Kirchhoff and Kiverstein hold that cultural practices are so important that they can be said, in a sense, to be vehicles of a subject's phenomenal consciousness. Their claim seems to be that cultural practices play a constitutive role in determining the *expected precision* of the incoming sensory signal. That is, they contribute to creating a system's expectation for certain very reliable streams of prediction error, which enable an agent to quickly deploy its own embodied skills to effectively cope with some relevant environmental contingency (Kirchhoff and Kiverstein 2019a: 94-100; 2020), thereby shaping one's subjective experience. Kirchhoff and Kiverstein provide a variety of examples of this, ranging from phoneme recognition (Roepstorff *et al*. 2010, cited in Kirchhoff and Kiverstein 2019a: 99-100) to color discrimination (Thierry *et al*. 2009, cited in Kirchhoff and Kiverstein 2019a: 98).

The example they discuss the most, however, is the culture shock Eva Hofmann felt when moving from Poland to Canada (Kirchhoff anf Kiverstein 2019a: 110-112; 2020). They argue that to account for the sense of estrangement Eva felt when moving, we need to refer to her active engagement with the cultural practices of her native cultural (Polish) environment, and her inability to enact such practices in the new (Canadian) cultural environment. This makes Eva's sensory states in her new cultural niche are surprisal-inducing, as they are not the ones she learned to predict. Her neural duplicate in the appropriate (Polish) environment would instead register no surprisal, courtesy of a compliant cultural environment (Kirchhof and Kiverstein 2019a: 113-114). Hence, in Kirchhoff and Kiverstein view, the cultural practices Eva took part in and learned while in Poland qualify as vehicles of her phenomenal experience. Importantly, Kirchhoff and Kiverstein (2020: 7-9) would add that these practices constitute Eva's phenomenology *diachronically*. This means, roughly, that the culture shock Eva felt is not exclusively constituted by factors present in the here-and-now (i.e. the moment when the culture shock is experienced). It is also partially constituted by the factors present in Eva's past that account for her experience now. The vehicles of Eva's experience, diachronic constitution suggest, are not just spread out in space. They are also spread out *in time*.

Crucially, Kirchhoff and Kiverstein argue that these culturally-leaden modifications of conscious experience are not due to the acquisition of specific neural representations. Rather, they are due to the constant agent-environment interaction; and cultural practices should be seen as elements regulating the behavior of the agent-environment coupled system, like macroscopic order parameters organizing the behavior of a dynamical system over very long timescales (Kirchhoff and Kiverstein 2020:7).[12] In their view, brains are *not* equipped with culturally learned representations; rather, they are nodes in a complex web of loopy causal relations, which are constantly transformed by the culturally regimented loops traversing them.

We end our exposition of Kirchhoff and Kiverstein's complex position here, owing to space limitations. We acknowledge that our summary is incomplete[13], and that it does not convey the entire depth of Kirchhoff and Kiverstein's overall position. But it is now time for us to articulate our claims.

## 4 - Extended consciousness in predictive processing: three problems

[12] Notice that, according to Kirchhoff and Kiverstein, this is exactly the role that the "wider" Markov Blanket should play in cases of dynamical entanglement.
[13] We have been silent, for instance, on Kirchhoff and Kiverstein's complex proposal of a diachronic account of constitution, and we have glossed over a variety of themes proposed in Kirchhoff and Kiverstein's book.

### 4.1 - Cultural practices do not seem to support consciousness vehicle externalism

As sketched above, Kirchhoff and Kiverstein (2019a; 2020) take cultural practices to be material vehicles of a subject's phenomenal consciousness: given the process ontology they assume (e. g. Kirchoff 2015), the vehicles of consciousness are not physical states but rather by physical *processes* extended in time. Among these processes, there are cultural practices, which determine (at least partially) the expected precision of incoming error signals. Hence, they play a role in error minimization, thereby partially determining a subject's phenomenally conscious states.

Yet, the account of expected precision PP offers is purely neural. PP suggests that neurons reporting prediction errors with a high expected precision have their post-synaptic gain increased, so as to make their prediction error dominate subsequent processing. Conversely, neurons reporting prediction errors with low expected precision get silenced by decreasing their post synaptic gain. Other mechanisms contribute to precision estimation too: dopamine can be used to quickly shift the post synaptic gain of error neurons, and synchronization of error reporting neurons might further increase their ability to influence cortical processing (see Feldman and Friston 2010; Friston 2012b; Friston *et al*. 2012). Collectively, these mechanisms allow our brains to "attend" only the most significant prediction error streams, thereby simplifying the solution of various cognitive tasks. Yet these mechanisms seem to be *neural* mechanisms. No *external* vehicle is present in the picture of precision estimation PP proposes. And given that the mechanisms for precision estimation that PP identifies operate at the neuronal level, it is hard to see *just how* cultural practices could be integrated in such a picture.[14]

[14] One could challenge this verdict, arguing that many external signs (e.g. traffic lights, stop signs, *etc*.) do play a role in settling our expectations about the precision of the incoming sensory signals, thereby functioning like a vehicle controlling our precision estimates (see Clark 2016, Ch. 8 and 9). We doubt, however, that Kirchhoff and Kiverstein can leverage such a challenge against us. This is because this case for vehicle externalism about

It could be argued that Kirchhoff and Kiverstein (2020: 7-9) provide an answer to this question, suggesting that cultural practices act as macroscopic order parameters on the agent-environment system. Yet, we see two problems with this suggestion. First, order parameters in dynamical systems typically represent physical magnitudes, such as the temperature gradient controlling the rolling motions of a heated liquid (e.g. Kelso 1995: 6-8), the frequency of operation of a central pattern generator (e.g. Kiebel et al. 2009) or the relative angle phase of two oscillators (e.g. Schmidt and Richardson 2008: 283). Hence, it is not clear whether the concept of an order parameter could be *rightfully* applied to cultural practices. Secondly, even allowing cultural practices to play a constraining role on neuronal dynamics similar to that of order parameters, it is hard to see *in what sense* such constraints would qualify as external vehicles of consciousness. Order parameters are not physical *components* of a dynamical system. They are mathematical values that perspicuously describe the state of a dynamical system in a dynamical model. Hence, even assuming that cultural practices can be cast as order parameters (or something analogous to them), they wouldn't be *components* of the coupled dynamical system they describe (i.e. a subject dynamically entangled with its environment). Thus, they wouldn't be vehicles inside that system.[15]

Perhaps focusing on a concrete case might clarify *how* cultural practices might function as external vehicles of subjective experience. Kirchhoff and Kiverstein (2019a: 98-99) take the experiment performed by Thierry *et al*. (2009) as evidence of the fact that cultural practices support CVE. Since the experiment is clearly described, we will consider it at length.

The study (Thierry *et al*. 2009) involved two groups of participants speaking two different native languages; namely English and Greek. Participants of both groups were required to

---

precision estimation is built upon a "first wave"-style parity argument for vehicle externalism (cfr. Constant *et al*. 2019a: 16-18). But Kirchhoff and Kiverstein (2019a, Ch. 1) *do not* endorse such a "first wave" form of vehicle externalism, and are in fact quite critical of it.

[15] Many thanks to **two colleagues** for discussion on these issues.

perform an odball stimulus discrimination. The stimulus was a sequence of squares, and the oddball was a circle. Participants had to press a button as soon as they noticed the circle. Notice that the stimuli were presented at a fixed rate, hence participants could not influence the way stimuli were presented. The neural activity of each participant was captured through an EEG cap. Crucially, albeit participants were required to discriminate *shapes*, stimuli could also vary in *color*. In total, four colors were used: light and deep blue and light and deep green. Variations in color, however, were *task irrelevant*: participants had to press the button only if the oddball (circular) stimulus appeared, independently from its color. Now, there is a crucial difference between English and Greek native speakers when it comes to the colors. Whereas both English and Greek use a single word for green (whether deep or light), only English uses a single word for blue. Greek uses two: *galázio* (light blue) and *ble* (deep blue).

Thierry and colleagues found that the early visual cortex of Greek (and only Greek) native speakers responded differently to the two task irrelevant shades of blue. Both English and Greek native speakers respond in the same way to the two shades of green. The researchers concluded that their data support the claim that color terminology can influence early visual processing.

We really do not see *how* this experiment is supposed to bolster the case for CVE. Let us start with consciousness. The experiment does *not* establish that Greek and English native speakers *experience* color differently. The data speaks only of a "[...] relationship between native language and *unconscious, preattentive color discrimination* rather than simply conscious, overt color categorization" (Thierry *et al*. 2009: 4568; *emphasis added*). Moreover, whether early visual cortices qualify as neural correlates of consciousness is still a matter of debate (Chalmers 2000; Blake *et al.* 2014; Koch *et al.* 2016), so it is at least in principle possible that color terminology only influences *non conscious* visual processing.

Now, suppose (for the sake of argument) that color terminology influences color phenomenology. Still, in the case at hand, it is not clear what are the relevant external vehicles that should "extend" a subject's color phenomenology. There are empirical studies that try to assess the impact of external objects on cognitive processing (e.g. Vallé-Tourangeau *et al.* 2016; Bocanegra *et al*. 2019), and these studies can be used to offer empirical support to vehicle externalism. But in these studies experimenters carefully describe the "external vehicles" involved, and what sort of effects they have on cognition. Thierry and colleagues do nothing of that sort. They neither describe the external vehicles that might "extend" their subject's consciousness, nor they try to describe the impact of such vehicles on subjective experience. And *rightfully so*: as described in their paper, their experimental procedure just does not involve any external vehicle.

Moreover, the experimental subjects do not appear to be dynamically entangled to anything in the environment. Participants were only required to press a button when the oddball was detected, and stimuli were presented every 800ms and flashed for 200ms, regardless of the subjects' responses. There is no closed causal or sensorimotor loop knitting together participants and environment in a single system. There just seems to be *no* instance of continuous reciprocal causation or coupling (Palermos 2014). And, in fact, *contra* dynamical views of cognition (Hurley 2001) the participant's task is easily decomposable in a linear sequence of input (stimulus reception) - cognition (discrimination) - action (eventual button pressing).

Kirchhoff and Kiverstein (2020: 5-7) might object that mastering a cultural practice is an *essentially* extended affair since a subject's mastery of a cultural practice *requires* that the subject constantly attunes to her cultural environment, interacting with it.[16] Hence, cultural

[16] Similar arguments for the essential extendedness of cognitive capacities are provided in (Noë 2004: 110-115) and (Hurley 2010: 138-143).

practices cannot be fully internalized in a set of culturally-shapened neural representations: mastering them requires constant agent-(cultural) environment interactions.

But this argument appears unconvincing. In general, the fact that an agent's mastery in a domain is achieved and maintained by means of constant environmental interactions does not pull the *vehicles* realizing that mastery out of the agent's body. Amy's mastery in weightlifting, for instance, has been achieved (and must constantly be maintained) by engaging with environmental props and tools (e.g. barbell weights). Yet, the material vehicles realizing her mastery (e.g. her muscle tone and her practical know-how on how to lift heavy weights) seem entirely internal to her body. And the same seems to hold for the mastery of cultural practices: one's mastery of a language depends critically on one's continued linguistic engagement with a certain linguistic environment. But the vehicles realizing one's linguistic mastery (e.g. one's Broca's area) sit squarely in one's skull.

Importantly, vehicle internalism *allows* cultural practices to have an effect on perceptual consciousness. The case of phonetic recognition (i.e. how our perception of linguistic stimuli changes based on the languages we know; see Roepstroff *et al*. 2010) persuasively exemplifies the effects of culture on our phenomenology. Indeed, the theoretical apparatus of PP seems almost ideally suited to account for such effects (see Lupyan *et al*. 2020 for a recent review). This is because PP heavily stresses the role of prior expectations (both about the incoming sensory inputs and their precision) in perception, and in particular in determining perceptual *content.* Hence, if cultural practices tailor a subject's estimate of precision, and these estimates impact (or concur to impact) a subject's perceptual content, then cultural practices can impact the *content* of a subject's consciousness.

Notice that a form of *content* externalism would be broadly consonant with Kirchhoff and Kiverstein's overall position on cultural practices. As noted by Clark and Chalmers (1998), content (as opposed to vehicle) externalism tends to stress factors lying in a *subject's past history,* such as the subject's embedding in a linguistic community (e.g. Putnam 1975). And these seem to be the factors that Kirchhoff and Kiverstein emphasize when discussing the case of Eva's culture shock and the diachronic constitution of her experience (see Kirchhoff and Kiverstein 2019a: 110-112; 2020: 9-10). Moreover, in at least some cases (e.g. Kirchhoff and Kiverstein 2019a: 111) they speak of cultural practices as settling our prior expectations. But, on the face of it, expectations are contentful states, whose *content* might be at least partially determined by some relevant cultural practice. For instance, a person born in a British English speaking cultural niche might expect to read "colour" instead of "color".

So, perhaps Kirchhoff and Kiverstein's emphasis on cultural practices can vindicate (or motivate) some form of externalism about perceptual *content*. However, content externalism does not entail CVE. To begin with, contents and vehicles should not be conflated; and, in fact, their conflation is in general fallacious (Dennett 1991; Hurley 1998). Secondly, content and vehicle externalism are logically independent, and do not entail each other (Hurley 2010; Rowlands 2020). The fact that conscious content is (perhaps) determined by extraneural factors does not, in and by itself, entail that the *vehicle* of said content is extraneural.

### 4.2 - The Markov Blankets of the conscious mind

In the PP literature, the issues concerning vehicle externalism are typically framed in terms of Markov Blankets. Given that everyone agrees that Markov Blankets are multiple and nested, the relevant issues concern how to determine *which* Markov Blanket delineates the boundary of a subject's mind (e.g. Hohwy 2016; 2017; Clark 2017).

Kirchhoff and Kiverstein propose a crisp answer: the relevant Markov Blanket identifying the vehicles of a subject's phenomenally conscious mind is the "wider" Markov Blanket surrounding "smaller" coupled Markov Blankets and acting as an order parameter on their synchronization (Kirchhoff and Kiverstein 2019a: 78-81). This is because the self-evidencing behavior of the "wider" Blanket constraints and determines the self-evidencing behavior of the smaller Blankets, forming "a prediction error minimizing whole" (Kirchhoff and Kiverstein 2019a: 81).

Here, we wish to argue that this way of identifying the Markov Blanket surrounding the vehicle of a subject's phenomenal consciousness leads to very unpalatable results, which also clash with Kirchhoff and Kiverstein's claim that only temporally thick generative models can support phenomenally conscious mental states.

To see why, consider first the relevant notion of coupling at play. In the literature on Markov Blankets, the relevant notion of coupling is typically operationalized in terms of generalized synchrony (e.g. Friston 2013; Bruineberg and Rietveld 2014; Bruineberg et al. 2018a; Kirchhoff and Kiverstein 2019a: 108-110; 2020: fn. 3; Palacios et al. 2019).[17] Generalized synchrony refers to a physical relation holding between a set of oscillators. If two (or more) oscillators interact (and influence each other's oscillations), they will eventually start to oscillate either in phase (i.e. with the same rate) or in anti-phase (i.e. with opposite rates). When one of these two conditions is reached, then generalized synchrony among the oscillators has been established, and the two oscillators can be considered as a single coupled system.

Secondly, consider the (seemingly well-established) claim that any two PP systems busy modelling each other rapidly fall into generalized synchrony (e.g. Friston and Frith 2015a; 2015b; Palacios *et al*. 2019). The reason as to why this is the case can be clearly expressed as

[17] This holds true also in some dynamical models (e.g. Schmidt and Richardson 2008; Chemero 2009: 85-98). Intriguingly, (Kirchhoff et al. 2018) seem to point to a different, more demanding, notion of coupling. However, as far as we can see, such a notion of coupling is still lacking an operationalization.

follows. Suppose that two agents (A and B) partake in a turn based activity, and suppose that they are both busy predicting what the other agent will do in its own turn. Now, when it's A's turn, B will try to predict the sensory consequences of A's moves. But, and this is the crucial point, *A is predicting them too*. For, if PP is correct, in order for A to take a move, A is bound to engage in active inference; and it is thus forced to predict the sensory consequences of its own moves. This is why the activity of two mutually predicting PP systems will tend to synchronize.

Different sources of empirical evidence reinforce the point, suggesting that interacting subjects tend to synchronize in a number of ways (see Wheatley *et al*. 2012; Coey *et al.* 2012; Tognoli *et al*. 2020 for reviews). Behavioral experiments suggest that interacting subjects spontaneously synchronize their limb movements (e.g. Schmidt *et al*. 1990; Richardson *et al*. 2005; see also Schmidt and Richardson 2008 for a review), postural sway (Shockley *et al*. 2003; Goodman *et al*. 2005; Richardson *et al*. 2007) and autonomic responses (Kang and Wheatley 2017). More recently, neuroimaging experiments deploying hyperscanning techniques have highlighted that when two subjects interact, their neuronal activity synchronizes in various regions of interest relevant for social cognition (e.g. Stephens *et al*. 2010; Liu *et al*. 2015; Jiang *et al*. 2015 Liu *et al*. 2016; see Valencia and Froese 2020 for a nice review), thereby providing strong evidence in favor of the claim that the neural activity of interacting subjects synchronizes. Hence, interacting subjects are coupled (according to the relevant notion of coupling at play here) at multiple levels.

Consider now the computational simulations presented in (Palacios *et al*. 2020) and (Friston *et al*. 2015). If correct, these simulations show that when two or more free-energy minimizing systems interact with each other, they naturally tend to form a wider system, with *its own* Markov Blanket, provided the interacting (i.e. "smaller") systems have *at least some* prior expectation in common.

But human subjects surely share at least some priors. Some of those are prior expectations regarding the statistics of the environment, to which all perceptual systems must be attuned (Orlandi 2014). For instance, humans expect natural light to illuminate objects from above and slightly on the left (e.g. Mamassian *et al*. 2002). Other shared priors regard our bodily dynamics, and the ways in which we can behave to achieve our goals (e.g.Kilner *et al*. 2007; 2011; Donnarumma *et al*. 2017). Consider further culturally established prior expectations. These practices are said to establish regimes of shared expectations among the members of a culture (Constant *et al*. 2019b; Kirchhoff and Kiverstein 2019a: Ch. 5; 2020). Taken together, all these threads of evidence strongly suggest that humans have at least some common priors. Hence, if the results of the simulations provided by Friston *et al.* (2015) and Palacios *et al*. (2020) are correct, the interaction of human subjects (which, if PP is correct, are free-energy minimizing systems) will naturally let a new system, with a "wider" Markov Blanket, emerge.

It thus appears that, anytime two subjects interact, they can be modeled as two synchronized Markov Blankets surrounded by a "wider" Blanket. And the internal dynamic of such a "wider" seems to act as an order parameter on the dynamics of the "smaller" Blankets it contains: after all, in many cases of synchronization, the order parameter is the relative phase of the two oscillators (e.g. Haken et al. 1985; Kelso 1995), which seems precisely what the "wider" Blanket enshrouds.

Notice that such a wider Blanket seems precisely the Markov Blanket that, according to Kirchhoff and Kiverstein, identifies the vehicles of a subject's phenomenal consciousness. Hence, if Kirchhoff and Kiverstein's Markov Blanket proposal on how to identify the vehicles of a subject's phenomenology is accepted, it seems that interacting subjects end up being vehicles of each other's phenomenology.

This would surely strike many as intuitively unappealing. But intuitively unappealing claims are not necessarily wrong. Thus, the fact that Kirchhoff and Kiverstein's defense of CVE ends up delivering intuitively unappealing results is not *necessarily* a problem for them.

What is a problem for them, however, is that such a "multi-subject" Markov Blanket does not identify a *temporally thick* model.[18] Yet, Kirchhoff and Kiverstein claim that only temporally thick models can be vehicles of phenomenal consciousness (Kirchhoff and Kiverstein 2019a: 53; 106-107). Hence, the way in which Kirchhoff and Kiverstein resort to Markov Blankets to individuate the boundaries of the conscious mind clashes with their theoretical commitments concerning the fact that only temporally thick models can give rise to subjective experience. So, either the way in which Kirchhoff and Kiverstein resort to Markov Blankets to identify the boundaries of a subject consciousness should be abandoned, or it is false that only temporally thick models can be vehicles of consciousness. At any rate, Kirchhoff and Kiverstein's position on the matter seems to lead to an inconsistency. *Something* must go.[19]

Why do we claim that the multi-subject model identified by the "wider" Blanket is not temporally thick? Recall the heuristic proposed by Friston (2018: 5-6): to determine the temporal thickness of a model it is sufficient to consider the time lapsed between successive "visits" to some particular states. The longer that time, the thicker the model. A bacterium, for instance, might revisit a state *x* every half an hour. Its model is thus significantly

---

[18] Recall that, in this context, "model" simply refers to the internal states of a Markov Blanket.

[19] One might object that Kirchhoff and Kiverstein resort to Markov Blankets only to formalize the boundaries of the mind, which need not coincide with the boundaries of the *conscious* mind. Yet, Kirchhoff and Kiverstein (2019a: 104) take the machinery PP describes to be the physical machinery of consciousness. And Markov Blankets presumably individuate the boundaries of *that* machinery. It is also worth noting that Kirchhoff and Kiverstein (2019a: 62-63) introduce Markov Blankets to deal with Chalmers's (2019) argument argument against CVE. It thus seems natural to think that, on Kirchhoff and Kiverstein's view, Markov Blankets should identify the boundaries of the conscious mind.

shallower than the model a typical human possesses, given that a typical human visits certain states only once a year (e.g. throwing a party for one's own birthday). A temporally thick generative model *loops* through its state space; and the temporal trajectory of that loop is an indicator of the temporal thickness of the model (or so Friston suggests). And if a model traces no loop in its state-space, then it is not temporally thick.

But, and we believe the following point is relatively uncontroversial, not all human interactions have the required loopy structure. Surely *some* human interactions have it: a group of friends can meet, say, once every year to commemorate some particular event. But many other human interactions are *one-shot* interactions. Consider, for instance, an applicant's interaction with the interviewer during a job interview. Both the interviewer and the applicant share many prior expectations, and both are prediction-error minimizing systems. So, if the simulations and the empirical evidence discussed above are correct, it seems that their interaction will tend to make them synchronize in various ways; and there will be a Markov Blanket "surrounding" them both. But the model identified through such a Blanket will not re-visit any of the states it happens to visit in the future - typically, applicants and interviewers do not periodically meet to re-enact job interviews. So, the model they jointly instantiate during the interview will not, according to the relevant heuristic Friston proposes, be temporally thick.

The same, it seems to us, holds true even if a more regimented notion of temporal thickness is deployed. For instance, temporal thickness is often understood in terms of the model's hierarchical structure, in which hierarchically higher levels predict (and postdict) the incoming inputs at a higher temporal scale (Tani 2016, Ch. 9-10; Friston *et al*. 2017). It is not at all clear whether multi-subject interactions have the required hierarchical structure, or in which sense two interacting subjects compose a functional hierarchy.

#### 4.3 - DEUTS and the consciousness bloat

One popular objection to vehicle externalism about cognition is the so-called "cognitive bloat" objection (Sprevak 2009; Rowlands 2010; 2020; Allen-Hermanson 2013). The objection basically points out that the criteria vehicle externalism offers to identify external vehicles are too easily satisfied, thereby implying that all sorts of things are vehicles of a subject's cognitive state. This is undesirable (if not a *reductio* of vehicle externalism) because it marrs the explanatory and scientific credential of vehicle externalism. An unruly amount of factors plays a causal-explanatory role in accounting for an agent's cognitive performance, and surely not all these factors are vehicles of cognition - and defenders of vehicle externalism do not wish to claim otherwise (Clark 2008: 76-81). Here, we claim that the DEUTS argument licenses a similar "bloat" worry when it comes to the vehicles of a subject's phenomenal consciousness.

To see why, recall the general structure of the DEUTS argument. The first step is a commitment to dynamicism: cognitive processes are often constituted by agent-environment sensorimotor interactions. The best way to explain these interactions is through the tools of dynamical systems theory. Once these tools are deployed, one is often forced to model the agent and the environment as a single coupled system, whose joint behavior accounts for the production of cognitive outputs. Hence, cognitive outputs are produced by an "extended" coupled system; just as cognitive vehicle externalism requires (Chemero 2009; Palermos 2014; Lamb and Chemero 2018; Kiverstein 2018).

The second step consists in showing that dynamically entangled subjects cannot be phenomenal duplicates *only* by being neural duplicates. If one closely scrutinizes *how* the experience of a dynamically entangled subject evolves over time, one notices that it is necessary, in order for a given temporally sequence of phenomenal states to be experienced, that agent and environment interact in a certain way; and thus that certain environmental features qualify as external vehicles of experience (Hurley 1998: Ch. 8; Kirchhoff and Kiverstein 2019: 36).

Crucially, this line of thought is supposed to provide a *discriminating* way to appeal to the causal spread of cognitive processing in order to establish CVE (Hurley 1998: 330). Every agent is a node in a massive causal network connecting it to an unruly manifold of environmental features (e.g. the oxygen an agent breathes), not all of which are putative external constituents of the agent's phenomenal machinery. To make her account discriminate between mere environmental causes and genuine external vehicles, Hurley (1998: 329-332) reasoned as follows: vehicles explain the obtaining of any particular mental state. The tokening of an appropriate vehicle is what in virtue of which each and every mental state obtains. Hence, what needs to obtain in order for a phenomenal state to obtain is the vehicle of that state. But not every environmental feature needs to obtain in order for a phenomenal state to obtain. So the argument *discriminates* between putative external constituents and environmental factors merely causally impacting upon a subject's consciousness.

We believe that Hurley's line of thought is not discriminatory enough. With enough ingenuity, one can always cook up a scenario in which the DEUTS argument ends up counting the oxygen in the atmosphere or the paint on the walls one is seeing (or whatever) as vehicles of a subject's phenomenal experience. Or so, at least, we contend.

Consider the following twist on Hurely's thought experiment. Suppose that a subject S and S's twin TS are in entirely white rooms, both with a ball at their right. Suppose that the layout of their environments are identical. The only difference is the amount of oxygen[20] present in

[20] We chose this example because Kirchhoff and Kiverstein (2020:12) clearly state that they do not wish to consider oxygen as a constituent of the phenomenal machinery.

the rooms: whereas S is in a normally oxygenated environment, TS isn't. In fact, there is no oxygen in TS's environment. Lastly, let both S and TS interact with the balls, so as to get dynamically entangled with the environment.

It seems clear that, given the proposed setup, the phenomenology experienced by S and TS, as well as their neural states, will diverge drastically. Whereas S will feel what it is like to play with a ball, TS will feel what it is like to choke to death. Whereas S's brain will instantiate rich patterns of activity, TS's brain will not (at least, not for long). Hence, we cannot duplicate S's temporally extended phenomenology in TS. But, as Hurley (1998: 331) writes in regard to mental states: "if not duplicable, then not the whole vehicle". Hence, the entire vehicle of S's phenomenology has not been duplicated in TS's scenario. But, by stipulation, the only difference between S and TS's environments is the presence of oxygen. And surely, were oxygen present in TS's room, TS *too* would feel what it is like to play with the ball: adding oxygen would thus allow us to duplicate S's experience in TS. But then the DEUTS argument forces us to conclude that oxygen is a constituent of S's phenomenal machinery, something that surely Kirchhoff and Kiverstein (2020: 12), as well as many other philosophers, wish to deny.

Similar unappealing results proliferate easily. Consider what would happen if the color of the walls in S and TS's rooms were different. Or if S and TS's balls had a different color. Or if TS were made to wear medieval armor. Or if the rooms had different light bulbs emitting differently colored light. It seems correct to say that, in all those cases, the temporally extended phenomenology of S (and the way in which S's neural states evolve through time) could not be duplicated in TS. And it seems equally correct to say that, in all those cases, eliminating the environmental feature acting as a difference-maker in between the two cases would allow S's phenomenology to be duplicated.

But if this is correct, then a systematic application of the DEUTS argument seemingly leads us to conclude that any (or almost any) environmental parameter contributing to the temporal unfolding of a dynamically entangled subject's experience is a *vehicle* of that experience. This, it seems to us, is an unwelcome result, for it suggests that the DEUTS argument leaves CVE open to a nasty "consciousness bloat" objection.

Perhaps the problem could be avoided by supplementing DEUTS with some further criteria enabling us to discriminate, among the candidate vehicles DEUTS identifies, *genuine* vehicles of consciousness from *spurious* ones. And, if we understand them correctly, Kirchhoff and Kiverstein (2019b; 2020) propose two such criteria. We examine them in turn.

The first criterion is proposed in Kirchhoff and Kiverstein (2019b: 16-18).[21] The idea seems to be the following: an external candidate vehicle of consciousness really qualifies as a vehicle only if it contributes to an agent's free-energy/prediction error minimization over time. As they write:

> "The self-evidencing nature of biological agents blocks the threat from cognitive bloat. External resources form a part of an agent's mind when they are poised to play a part in the processes of active inference that keep surprise to a minimum over time (i.e. that minimise free-energy)." (Kirchhoff and Kiverstein 2019b: 17).

We do not think that this effectively blocks the threat bloat argument poses. To see why, consider *interoceptive* active inference. If PP is correct, brains are busy predicting the incoming input *in all modalities*. Importantly, this means that brains will not just try to guess the incoming sensory signal in the *exteroceptive* modalities (bluntly put, the five senses traditionally understood); they will also try to guess the *proprioceptive* signals (i.e. the "sense" of kinesthesia and self-movement) and the *interoceptive* signals (i.e. the "sense" of

[21] To be fair, in that paper Kirchhoff and Kiverstein *do not* detal with CVE directly, but only with *cognitive* vehicle externalism. However, since Kirchhoff and Kiverstein (2019a: 104) take the PP explanations of *cognitive* phenomena to be also an account of *phenomenal consciousness*, the argument they offer to block the *cognitive* bloat should also block the *consciousness* bloat.

one *internal* bodily state). Albeit typically associated with emotional responses (e.g. Seth 2013; Pezzulo 2014; Seth and Friston 2016) prediction of interoceptive signals is functionally on a par with the prediction of exteroceptive and proprioceptive signals.[22] There are thus two general ways to minimize error relative to interoceptive predictions: changing the predictions to make them fit the incoming interoceptive signal *or* changing the incoming interoceptive signal to make it fit the predictions. The latters is *interoceptive* active inference.

A concrete example of interoceptive active inference occurs when humans try to reduce the error relative to the prediction of their bodily temperature being around 36.6° (Bruineberg 2017; 3). One way to reduce that prediction error is to wear appropriate clothes, given the environmental temperature (e.g. wearing a sweater during the winter). But, most times, humans *do* wear appropriate clothes, given the environmental temperature. So it seems that clothes are well poised to play the desired error minimizing role over time, just as Kirchhoff and Kiverstein ask. It thus seems correct to say that *clothes* are part of the machinery that minimizes (interoceptive) prediction error on average and in the long run. And given that Kirchhoff and Kivertein take that machinery to be the machinery of consciousness (Kirchhoff and Kiverstein 2019a: 104; 2020), it seems correct to conclude that, on their account, clothes are cogs in our phenomenal machinery. This, to us, seems sufficient to conclude that the consciousness bloat objection is not avoided.

Notice, further, that focusing on the external vehicles that minimize surprise *overtime* seems to clash with Clark's (2017) metamorphosis argument, upon which Kirchhoff and Kiverstein rely. The reason is this: presumably, the question whether some candidate vehicle plays an error/free-energy minimizing role *overtime* has a crisp yes or no answer. Either something plays such a role often enough, or it doesn't. Hence, if the suggestion put forth by

[22] So much so, that there can be interoceptive perceptual illusions; see (Iodice *et al*. 2019) for a nice example.

Kirchhoff and Kiverstein (2019b) is correct, an agent's Markov Blanket is somewhat rigidly fixed. Yet, the metamorphosis argument stresses the *plasticity* of an agent's Markov Blanket, suggesting that it can easily be altered.

Consider now the second anti-bloat criterion, proposed in Kirchhoff and Kiverstein (2020: 11-12). The criterion revolves around *counterfactual manipulations*. To understand it, recall first that free-energy (which, under simplificatory assumptions, corresponds to prediction error) is the sum of two quantities: the *surprisal* of a sensory state and the $D_{KL}$; where the $D_{KL}$ measures of how much the probability distributions encoded in an agent's expectations differ from the *actual* probability distributions defined over environmental causes.

Provided this, Kirchhoff and Kiverstein suggest that we should identify as external vehicles of a subject's consciousness only the elements upon which a counterfactual intervention would change the subject's $D_{KL}$ and thus the subject's phenomenology. In their view, this simple test is sufficient to tell apart the external factors which are part of a subject's phenomenal machinery from the ones that merely *causally interact* with that machinery (Kirchhoff and Kiverstein 2020: 12).[23]

We must confess that we do not see how this simple test can help Kirchhoff and Kiverstein's cause, as changes in the $D_{KL}$ are typically associated with perceptual inference. If a subject correctly determines the external cause of the incoming sensory signals, the subject's $D_{KL}$ will be small. Conversely, if a subject mis-infers the cause of the incoming signal, the subject's $D_{KL}$ will be sizable. If this is correct, it is obvious that *any* intervention on an external cause of the sensory signal will change a subject's $D_{KL}$, and presumably the

[23] Kaplan's (2012) mutual manipulability criterion is another criterion that relies on counterfactual interventions to tell apart the genuine constituents of a subject's mental machinery from factors that merely causally impacing it. Importantly, however, Kaplan's criterion requires at least two counterfactual interventions: a "bottom-up" intervention on the putative vehicle and a "top-down" intervention on the relevant (allegedly extended) cognitive phenomenon. Kirchhoff and Kiverstein, in contrast, seem only to require "bottom up" interventions on the putative vehicle. So, Kirchhoff and Kiverstein's criterion is distinct from Kaplan's.

subject's phenomenology.

Notice that such interventions will alter the $D_{KL}$ of active and *passive* perceivers alike: even if a subject is "perfectly still", intervening on the environmental causes of their sensory inputs will alter the subject's $D_{KL}$. Thus, the criterion proposed by Kirchhoff and Kiverstein (2020) seems even *less* discriminating than the original DEUTS argument. For, according to the DEUTS argument, something qualifies as an external vehicle of consciousness *only if* a subject is dynamically coupled to it. But this requirement is not present in Kirchhoff and Kiverstein's (2020) criterion. Hence, it is more liberal than the original DEUTS argument, and thus cannot be used to solve the liberality problems afflicting DEUTS.

We thus conclude Kirchhoff and Kiverstein additional criteria do not succeed in rescuing the DEUTS argument.

**5 - Conclusion**

In this paper, we have examined some aspects of Kirchhoff and Kiverstein's DEUTS-based marriage of PP and CVE, arguing that Kirchhoff and Kiverstein's position is susceptible to a nasty "consciousness bloat" objection, and that their emphasis on cultural practices does not contribute to establishing the truth of CVE.

Importantly, if the arguments we have presented here are correct, then, contrary to a popular opinion (e.g. Clark 2009; 2013; Kirchhoff and Kiverstein 2019a), DEUTS is not a strong argument in favor of CVE, as it is extremely susceptible to the "consciousness bloat" objection, which is inbuilt in its argumentative fabric.

Notice that we are *not* implying that CVE is false. There are other arguments in favor of CVE (e.g. Vold 2015) which might succeed where DEUTS fails. But, thus far, they have received very little attention: DEUTS has always been the focus of the debate on CVE. We thus suggest that time is ripe to put DEUTS in retirement, and find some new argument in favor of CVE.

**References**


Adams, R. A., *et al*. (2013). Predictions, not commands. Active inference in the motor cortex. *Brain Structure and Function*, 218(3), 611-643.

Allen-Hermanson, S. (2013). Superdupersizing the mind: extended cognition and the persistence of cognitive bloat. *Philosophical Studies*, *164*(3), 791-806.

Andrews, M. (2021). The math is not the territory. *Biology and Philosophy*, *36*(6), 1-19.

Baltieri, M., & Buckley, C. L. (2019). Generative models as parsimonious descriptions of sensorimotor loops. *Behavioral and Brain Sciences*, 42, https://doi.org/10.1017/S0140525X19001353.

Blake, R., *et al*. (2014). Can binocular rivalry reveal the neural correlates of consciousness?. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 369(1641), 20130211.

Bocanegra, B. R. (2019). Intelligent problem solvers externalize cognitive operations. *Nature Human Behavior*, 3(2), 136-142.

Bruineberg, J., & Rietveld E. (2014). Self-organization, free-energy minimization, and an optimal grip on a field of affordances. *Frontiers in Human Neuroscience*, *8*: 599.

Bruineberg, J. (2017). Active inference and the primacy of the "I can". In T. Metzinger, W. Wiese (Eds.), *Philosophy and Predictive Processing*: 5. Frankfurt am Main, The MIND Group. https://doi.org/10.15502/9783958573062.

Bruineberg, J., *et al*. (2018a). The anticipating brain is not a scientist: the free-energy principle from an ecological-enactive perspective. *Synthese*, 195(6), 2417-2444.

Bruineberg, J. *et al.* (2018b). free-energy minimization in joint agent environment systems: a niche construction perspective. *Journal of Theoretical Biology*, 455, 161-178.

Bruineberg, J. *et al*. (2020). The emperor's new Markov Blankets. *Preprint*, http://philsci-archive.pitt.edu/18467/

Buckley, C. L., *et al.* (2017). The free-energy principle for action and perception: a mathematical review. *Journal of Mathematical Psychology*, 81, 55-79.

Chalmers, D. J. (2000). What is a neural correlate of consciousness?. In T. Metzinger (Ed.), *Neural Correlates of Consciousness*, (pp. 17-39). Cambridge, MA.: The MIT Press.

Chalmers (2019). Extended cognition and extended consciousness. In M. Colombo, E. Irvine, M. Stapleton (Eds.), *Andy Clark and His Critics* (pp. 7-20). New York: Oxford University Press.

Chemero, A. (2009). *Radical Embodied Cognitive Science*. Cambridge, MA.: The MIT Press.

Clark, A. (2009). Spreading the joy? Why the machinery of consciousness is (probably) still in the head. *Mind*, 118(472), 963-993.

Clark, A. (2012). Dreaming the whole cat. *Mind*, 121(483), 753-771.

Clark, A. (2013). *Mindware* (2nd Edition). New York: Oxford University Press.

Clark, A. (2016). *Surfing Uncertainty*. New York: Oxford University press.

Clark, A. (2017). How to knit your Markov Blanket. In T. Metzinger, W. Wiese (Eds.), *Philosophy and Predictive Processing*: 3. Frankfurt am Main, The MIND Group. https.//doi.org/10.15502/9783958573031.

Clark, A., & Chalmers, D. J. (1998). The extended mind. *Analysis*, 58(1), 7-19.

Coey, C. A., *et al*. (2012). Coordination dynamics in socially situated nervous systems. *Frontiers in Human Neuroscience*, *6*: 164.

Constant, A., *et al*. (2019a). Extended active inference: constructing predictive cognition beyond skulls. *Mind and Language*, https://doi.org/10.1111/mila.12330

Constant, A. *et al*. (2019b). Regimes of expectations: an active inference model of social conformity and human decision making. *Frontiers in Psychology*, 10:679.

Corcoran, A., Pezzulo, G., & Hohwy, J. (2020). From allostatic agents to counterfactual cognizers: active inference, biological regulation, and the origins of cognition. *Biology and Philosophy*, *35*(3), 1-45.

Donnarumma, F., *et al*. (2017). Action understanding as hypothesis testing. *Cortex*, *89*, 45-60.

Dennett, D. (1991). *Consciousness Explained*. New York: Little Brown.

Fabry, R. E. (2017). Transcending the evidentiary boundary: prediction error minimization, embodied interaction, and explanatory pluralism. *Philosophical Psychology*, 30(4), 395-414.

Feldman, H., & Friston, K. (2010). Attention, uncertainty and free-energy. *Frontiers in Human Neuroscience*, *4*: 215.

Friston, K. (2005). A theory of cortical responses. *Philosophical Transaction of the Royal Society B: Biological Sciences*, 360(1456), 815-836.

Friston, K. (2009). The free-energy principle: a rough guide to the brain?. *Trends in Cognitive Sciences*, 13(7), 293-301.

Friston, K. (2011). What is optimal about optimal motor control?. *Neuron*, 72(3), 488-498.

Friston, K. (2012a). A free-energy principle for biological systems. *Entropy*, 14(11), 2100-2121.

Friston, K. (2012b). Predictive coding, precision and synchrony. *Cognitive Neuroscience*, *3* (3-4), 238-239.

Friston, K. (2013). Life as we know it. *Journal of the Royal Society Interface*, 10(86): 20130475.

Friston, K. (2018). Am I self-conscious? (Or: does self-organization entail self-consciousness?). *Frontiers in Psychology*, 9: 579.

Friston, K. (2019). Beyond the desert landscape. In M. Colombo, E. Irvine, M. Stapleton (Eds.), *Andy Clark and His Critics*, (pp. 174-190). New York: Oxford University Press.

Friston, K., & Frith, C. (2015a). A duet for one. *Consciousness and Cognition*, 36, 390-405.

Friston, K., & Frith, C. (2015b). Active inference, communication and hermeneutics. *Cortex*, 68, 129-163.

Friston, K., *et al.* (2012). Perception as hypothesis, saccades as experiments. *Frontiers in Psychology*, 3: 151.

Friston, K. *et al*. (2017). Deep temporal models and active inference. *Neuroscience and Biobehavioral Reviews*, *77*, 388-402.

Friston, K. *et al*. (2020). Sentience and the origin of consciousness: from Cartesian duality to Markovian monism. *Entropy*, *22*(5): 516.

Friston, K*., et al*. (2015). Knowing one's place: a free-energy approach to pattern regulation. *Journal of the Royal Society Interface*, 12(105), 20141383.

Friston, K., & Stephan, K. (2007). free-energy and the brain. *Synthese*, 159(3), 417-458.

Goodman, J. R. L., *et al*. (2005). The interpersonal phase entrainment of rocking chair movements. In H. Heft, K. L. Marsh (Eds.), *Studies in Perception and Action VIII: Thirteenth International Conference on Perception and Action* (pp. 49-53). Erlbaum.

Haken, H, *et al*. (1985). A theoretical model of phase transitions in human hand movements.

*Biological Cybernetics*, *51*(5), 347-356.

Hobson, J. A., & Friston, K. (2014). Waking and dreaming consciousness: neurobiological and functional considerations. *Progress in Neurobiology*, 98(1), 82-98.

Hohwy, J. (2013). *The Predictive Mind*. New York: Oxford University Press.

Hohwy, J. (2016). The self evidencing brain. *Noûs*, 50(2), 259-282.

Hohwy, J. (2017). How to entrain your evil demon. In T. Metzinger, W. Wiese (Eds.), *Philosophy and Predictive Processing*, 2, Frankfurt am Main: The MIND Group, https://doi.org/10.15502/9783958573048.

Hohwy, J. (2020). Self-supervision, normativity and the free-energy principle. *Synthese*, https://doi.org/10.1007/s11229-020-02622-2.

Hurley, S. (1998). *Consciousness in Action*. Cambridge, MA.: Harvard University Press.

Hurley, S. (2001). Perception and action: alternative views. *Synthese*, 129(1), 3-40.

Hurley, S. (2010). The varieties of externalism. In R. Menary (ed.), *The Extended Mind*, (pp.101-153). Cambridge, MA.: The MIT Press.

Hurley, S., & Noe, A. (2003). Neural plasticity and consciousness. *Biology and Philosophy*, *18*(1), 131 - 168.

Iodice, P., *et al*. (2019). An interoceptive illusion of effort induced by false heart-rate feedback. *Proceedings of the National Academy of Sciences*, 116(28), 13897-13902.

Jiang, J., *et al.* (2015). Leader emergence through interpersonal neuronal synchronization. *Proceedings of the National Academy of Sciences*, *112*(14), 4274-4279.

Kang, O., & Wheatley, T. (2017). Pupil dilatation patterns spontaneously synchronize across individuals during shared attention. *Journal of Experimental Psychology: General*, 146(4), 569-576.

Kaplan, D. M. (2012). How to demarcate the boundaries of cognition. *Biology and Philosophy*, 27(4), 545-570.

Kelso, S. (1995). *Dynamic Patterns*. Cambridge, MA.: The Mit Press.

Kiebel, S. J., *et al*. (2009). Recognizing sequences of sequences. *PLoS Comput. Biol.*, *5*(8): e1000464.

Kiefer, A., & Hohwy, J. (2018). Content and misrepresentation in hierarchical generative models. *Synthese*, *195*(6), 2387-2415.

Kilner, J. M., *et al*. (2007). Predictive coding: an account of the mirror neuron system. *Cognitive Processing*, 8(3), 159-166.

Kilner, J. M., *et al*. (2011). Action understanding and active inference. *Biological Cybernetics*, 104(1-2), 137-160.

Kirchhoff, M. D. (2015). Experiential fantasies, prediction, and enactive minds. *Journal of Consciousness Studies*. *22*(3-4), 68-92.

Kirchhoff, M. D. (2018). Predictive processing, perceiving and imagining: is to perceive to imagine, or something close to it?. *Philosophical Studies*, *175*(3), 751-767.

Kirchhoff, M. D., & Kiverstein, J. (2019a). *Extended Consciousness and Predictive Processing*. New York: Routledge.

Kirchhoff, M. D., & Kiverstein, J. (2019b). How to determine the boundaries of the mind: a Markov blanket proposal. *Synthese*, https://doi.org/10.1007/s11229-019-02370-y.

Kirchhoff, M. D., & Kiverstein, J. (2020). Attuning to the world: the diachronic constitution of the extended conscious mind. *Frontiers in Psychology*, https://doi.org/10.3389/fpsyg.2020.01966

Kirchhoff, M. D., *et al*. (2018). The Markov Blankets of life: autonomy, active inference and the free-energy principle. *Journal of the Royal Society Interface*, 15(138): 20170792.

Kiverstein, J. (2018). Extended cognition. In A. Newen, L. de Bruin, S. Gallagher (Eds.), *The Oxford Handbook of 4E Cognition*, (pp. 19-40). New York: Oxford University Press.

Kiverstein, J., & Farina, M. (2012). Do sensory substitution devices extend the conscious mind?. In F. Paglieri (ed.), *Consciousness in Interaction* (pp. 19-40). Amsterdam/Philadelphia: John Benjamins Publishing Company.

Koch, C., *et al*. (2016). Neural correlates of consciousness: progress and problems. *Nature Reviews Neuroscience*, 17(5), 307-321.

Laflaquiere, A. (2017). Grounding the experience of a visual field through sensorimotor contingencies. *Neurocomputing*, *268*, 142-152.

Lamb, M., & Chemero, A. (2018). Interaction in the open: where dynamical systems become extended and embodied. In A. Newen, L. de Bruin, S. Gallagher (Eds.), *The Oxford Handbook of 4E Cognition*, (pp. 147-162). New York: Oxford University Press.

Leinweber, M., *et al*. (2017). A sensorimotor circuit in the mouse cortex for visual flow prediction. *Neuron*, *95*(6), 1420-1432.

Liu, T. *et al*. (2015). Role of the right inferior frontal gyrus in turn-based cooperation and competition: a near-infrared spectroscopy study. *Brain and Cognition*, *99*, 17-23.

Liu, N. *et al*. (2016). NIRS-based hyperscanning reveals inter-brain neural synchronization during cooperative Jenga game with face-to-face communication. *Frontiers in Human Neuroscience*, *10*: 82.

Lupyan, G. *et al.* (2020). Effects of language on visual perception. *Trends in Cognitive Sciences*, https://doi.org/10.1016/j.tics.2020.08.005.

Mamassian, P., *et al*. (2002). Bayesian modelling of visual perception. In R. Rao, B. Olshausen, M. Lewicki, *Probabilistic Models of the Brain: Perception and Neural Functioning*, (pp.13-36). Cambridge, MA.: The MIT Press.

Maye, A., & Engel, A. K. (2013). Extending sensorimotor contingency theory: prediction, planning, and action generation. *Adaptive Behavior*, 21(6), 423-436.

Menary, R., & Gillett, A. (2020). Are markov blankets real and does it matter?. In D. Mendoça, M. Curado, S. Gouveia (Eds.), *The Philosophy and Science of Predictive Processing* (pp. 39-58). London: Bloomsbury Academic.

Millidge, B., *et al.* (2021). Whence expected free-energy?, *Neural Computation*, *33*(2), 447-482.

Noë, A. (2004). *Action in Perception*. Cambridge, MA.: The MIT Press.

Noë, A. (2009). *Out of Our Heads*. Basingstoke: Macmillan.

O'Regan, J. K. (2011). *Why Red doesn't Sound Like a Bell: Understanding the Feeling of Consciousness*. New York: Oxford University Press.

O'Regan, J. K., & Noë, A. (2001a). A sensorimotor account of vision and visual consciousness. *Behavioral and Brain Sciences*, 24(5), 939-973.

O'Regan, J. K., & Noë, A. (2001b). What it is like to see. A sensorimotor theory of perceptual experience. *Synthese*, 192(1), 79-103.

Orlandi, N. (2014). *The Innocent Eye. Why Vision is not a Cognitive Process*. New York: Oxford University Press.

Palacios, E. E., *et al*. (2019). The emergence of synchrony in networks of mutually inferring neurons. *Scientific Reports*, 9(1), 1-14.

Palacios, E. E., *et al*. (2020). On Markov Blanket and hierarchical self-organization. *Journal of Theoretical Biology*, 486:110089.

Palermos, O. (2014). Loops constitution, and cognitive extension. *Cognitive Systems*

*Research*, 27, 25-41.

Pepper, K. (2014). Do sensorimotor dynamics extend the conscious mind?. *Adaptive Behavior*, *22*(2) 99-108.

Pezzulo, G. (2014). Why do you fear the bogeyman? An embodied predictive coding model of perceptual inference. *Cognitive, Affective, and Behavioral Neurosciences*, 14(3), 902-911.

Pezzulo, G. *et al*. (2017). Model-based approaches to active perception and control. *Entropy*, 19(6): 266.

Putnam, H. (1975). The meaning of "meaning". *Philosophical Papers Vol. II: Mind, Language and Reality.* Cambridge: Cambridge University Press.

Ramstead, M. J. D., *et al*. (2020). A tale of two densities: active inference in enactive inference. *Adaptive Behavior*. *28*(4), 225-239.

Rao, R., & Ballard D. (1999). Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive field effects. *Nature Neuroscience*, 2(1), 79-87.

Roepstorff, A., *et al*. (2010). Enculturating brains through patterned practices. *Neural Networks*, 23, 1051-1059.

Rowlands, M. (2020). Externalism about the mind. In E. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (winter 2020 edition), https://plato.stanford.edu/archives/win2020/entries/content-externalism

Richardson, M. J., *et al*. (2005). Effects of visual and verbal interaction on unintentional interpersonal coordination. *Journal of Experimental Psychology: Human Perception and Performance*, 31(1), 62-79.

Richardson, M. J., *et al*. (2007). Rocking together: dynamics of intentional and unintentional social coordination. *Human Movement Science*, *26*, 867-891.

Rietveld, E., & Kiverstein, J. (2014). A rich landscape of affordances. *Ecological Psychology*, *26*(4), 325-352.

Rietveld, E., *et al*. (2018). Ecological-enactive cognition as engaging with a field of relevant affordances: the Skilled-Intentionality Framework (SIF). In A. Newen, L. de Bruin, S. Gallagher (Eds.), *The Oxford Handbook of 4E Cognition* (pp. 41-70). New York: Oxford University Press.

Schmidt, R. C., *et al*. (1990). Phase transitions and critical fluctuations in the visual coordination of rhythmic movements between people. *Journal of Experimental Psychology: Human Perception and Performance*, 16(2), 227-247.

Schmidt, R. C., & Richardson, M. J. (2008). Dynamics of interpersonal coordination. In A. Fuchs, V. K. Jirsa (Eds.). *Coordination: Neural, Behavioral and Social Dynamics*, (pp. 281-308), Springer, Berlin-Heidelberg.

Seth. A. K. (2013). Interoceptive inference, emotion, and the embodied self. *Trends in Cognitive Sciences*, 17(11), 565-573.

Seth, A. K. (2014). A predictive processing theory of sensorimotor contingencies: explaining the puzzle of perceptual presence and its absence in synesthesia. *Cognitive Neuroscience*, 5(2), 97-188.

Seth, A. K., & Friston, K. (2016). Active interoceptive inference and the emotional brain. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 371(1708), 2016007.

Seth, A. K., & Hohwy, J. (2020). Predictive processing as a systematic basis for identifying the neural correlates of consciousness. *Philosophy and the Mind Sciences*, *1*(2), https://doi.org/10.33735/phimisci.2020.II.64

Shockley, K. *et al*. (2003). Mutual interpersonal postural constraints are involved in cooperative conversation. *Journal of Experimental Psychology: Human Perception and Performance*, 29(2), 326-332.

Sprevak, M. (2009). Extended cognition and functionalism. *The Journal of Philosophy*, 106(9), 503-527.

Stephens, G. J., *et al*. (2010). Speaker-listener neural coupling underlies successful communication. *Proceedings of the National Academy of Sciences*, 107(32), 14425-14430.

Tani, J. (2016). *Exploring Robotic Minds*. New York: Oxford University Press.

Thierry, G. *et al*. (2009). Unconscious effects of language specific terminology on preattentive color perception. *Proceedings of the National Academy of Sciences*, 106(11), 4567-4570.

Tognoli, E., *et al*. (2020). Coordination dynamics: a foundation for understanding social behavior. *Frontiers in Human Neuroscience*, 14, https://doi.org/10.3389/fnhum.2020.00317.

Tschantz, A., *et al*. (2020). Learning action-oriented models through active inference. *PLoS Computational Biology*, *16*(4), e1007805.

Valencia, A. L., & Froese T. (2020). What binds us? Interbrain neural synchronization and its implications for theories of human consciousness. *Neuroscience of Consciousness*, 2020(1): niaa010.

Vallé-Tourangeau, F. *et al*. (2016). Insights with hands and things. *Acta Psychologica*, 170, 195-205.

Vázquez, M. J. C. (2020). A match made in heaven: predictive approaches to (an unorthodox) sensorimotor enactivism. *Phenomenology and the Cognitive Sciences*, *19*, 653-684.

Vold, K. (2015). the parity argument for extended consciousness. *Journal of Consciousness Studies*, *22* (3-4), 16-33.

Wheatley, T., *et al*. (2012). From mind perception to mental connection: synchrony as a mechanism for social understanding. *Social and Personality Psychology Compass*, 6(8), 589-606.

Wiese, W. (2018). *Experienced Wholeness*. Cambridge, MA.: The MIT Press.

Wiese, W., & Metzinger, T. (2017). Vanilla PP for philosophers. In T. Metzinger, W. Wiese (Eds.), *Philosophy and Predictive Processing*: 1. Frankfurt am Main, The MIND Group. https://doi.org/10.15502/9783958573024.