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

Consciousness vehicle externalism (CVE) is the claim that the material machinery of a subject's phenomenology partially leaks outside a subject's brain, encompassing bodily and environmental structures. The DEUTS argument is the most prominent argument for CVE in the sensorimotor enactivists' arsenal. In a recent series of publications, Kirchhoff and Kiverstein have deployed such an argument to claim that a prominent view of neural processing, namely predictive processing, is fully compatible with CVE. Indeed, in Kirchhoff and Kiverstein's view, a proper understanding of predictive processing mandates CVE. In this essay, we critically examine Kirchhoff and Kiverstein's argument. Our aim is to argue in favor of the following three points. First, that Kirchhoff and Kiverstein's emphasis on cultural practices lends no support to CVE: at best, it vindicates some form of content externalism about phenomenal content. Secondly, the criteria Kirchhoff and Kiverstein propose to identify a subject's phenomenal machinery greatly overgeneralize, leaving them open to a "consciousness bloat" objection, which is an analog of the cognitive bloat objection against the extended mind. Lastly, we will argue that the "consciousness bloat" problem is inbuilt in the very argumentative structure of the DEUTS argument. We will thus conclude that, contrary to the philosophical mainstream, DEUTS is not the best argument for CVE in the sensorimotor enactivists' argumentative arsenal.

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

1 - Introduction

The extended mind thesis, or *vehicle externalism*, is the claim that, at least sometimes, the material vehicles of mental processes are physically located outside a subject's biological shell (Hurley 2010; Clark 2013: 192-193). In this view, the physical machinery of the mind is, at least sometimes, distributed between brains, active bodies, and worldly props. Importantly, the scope of vehicle externalism is typically restricted to sub-personal cognitive processing and personal, but dispositional, mental states. (Clark and Chalmers 1998; Clark 2008).

Sensorimotor enactivism goes a step further, endorsing vehicle externalism about phenomenally conscious states (CVE), in particular perceptual ones.[1] According to sensorimotor enactivists, perception is something an *embodied agent achieves through sensorimotor interactions with its environment*, such as saccading over a target to explore its visual features, or by squeezing a sponge to feel its softness (Noë 2004, 2009). In this view, the *sensorimotor contingencies* (i.e. law-like linkages between bodily movements and changes of sensory stimulation) determine the phenomenal qualities of experiences (see Hurley and Noë 2003; O'Regan 2011, 2012), which, in turn, are brought about, or enacted, by the agent's sensorimotor interactions. Importantly, these interactions enact phenomenal qualities insofar as they are guided by the agent's *sensorimotor mastery*, that is, the agent's practical knowledge of sensorimotor contingencies. Hence, on the view sensorimotor enactivists offer, the phenomenal machinery of perception includes, as constituent parts, agent-environment interactions, and not just the neural processes enabling them (O'Regan

[1] Some linguistic stipulations: in the following, the unqualified term "consciousness" will be used as a synonym of phenomenal consciousness. Similarly, the unqualified term "perception" will refer to phenomenally rich perception. We will also use the phrase "phenomenal machinery" and "phenomenal circutery" to refer collectively to the material vehicles of conscious states. Notice further that, in this paper, we will only consider the sensorimotor enactivists' version of consciousness vehicle externalism, and "CVE" will only refer to it. Our arguments do not impact other forms of CVE, such as the one presented in (Vold 215).

and Noë 2001a; 2001b; Kiverstein and Farina 2012; Ward 2012; Pepper 2014).

CVE strikes many as an implausible, if not unpalatable, thesis; and it has thoroughly been attacked in the philosophical literature (Horgan and Kriegel 2008; Clark 2009; Chalmers 2019). As if philosophical counterarguments weren't enough, CVE is also put under significant pressure by the neurocomputational framework of predictive processing (PP). This is because PP seems able to re-cast sensorimotor contingencies in purely *neural* terms, as part of the knowledge encoded in a neurally realized generative model (Clark 2012; Seth 2014; Pezzulo et al. 2017; Baltieri and Buckley 2019). By doing so, PP poses an important threat to CVE.[2] For, if an agent's mastery of sensorimotor contingencies is a purely neural affair, and if sensorimotor contingencies only mediate *neural* activities, then there seems to be just *no* reason to endorse CVE.

In a series of recent publications, Kirchhoff and Kiverstein (2019a; 2019b; 2020) have set off to revise this dialectical situation. Their aim is to show that PP and CVE are compatible, if not to show that, once properly understood, PP *mandates* CVE.[3] 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 DEUTS argument (see Clark 2009; 2013: 222-225; Kirchhoff and Kiverstein 2019a: Ch. 3).

Our aim here is to evaluate Kirchhoff and Kiverstein's endeavor. We will argue that their proposal, as a whole, fails because of three distinct reasons; namely:

(1) Kirchhoff and Kiverstein's emphasis on cultural practices offers no support to CVE.

(2) Their way to identify the machinery of consciousness leads to a "consciousness bloat" problem, formally identical to the "cognitive bloat" problem plaguing the extended

[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).

[3] In all fairness, Kirchhoff and Kiverstein never state the point explicitly. Yet, this is entailed by their position. In fact, they hold that, properly understood, the mechanism responsible for PP is always realized widely by an agent-environment coupled system (see Kirchhoff and Kiverstein 2019a: Ch. 3, see also Kirchhoff 2015). They also hold that that mechanism just is the relevant phenomenal machinery (e.g. Kirchhoff and Kiverstein 2019a: 104). This straightforwardly entails that a properly understood PP is committed to CVE, and that if PP is correct, then CVE must also be correct.

mind

(3) The "consciousness bloat" is inbuilt in DEUTS, as the argument is structurally incapable of discriminating background causes impacting the phenomenal machinery from candidate external constituents

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, the DEUTS argument is *not* the best argument to secure CVE.

We'll work as follows: 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 conclusion will then follow.

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

PP is a neurocomputational framework providing a process theory to Friston's free energy principle (Friston and Stephan 2007; Hohwy 2020). Here, we provide an hyper-concise introduction to both, starting from the latter.[4]

The free energy principle states that biological self-organization obeys a simple imperative; namely that of minimizing *surprisal* - an information theoretic quantity roughly measuring the unexpectedness of sensory states, given the sensory states that an organism should expect to occupy to keep itself within its own biological bounds of viability (Friston 2012a; 2013; 2019).

Organisms, however, cannot quantify surprisal directly. Yet, they can evaluate 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 the system's

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

"beliefs" on 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[5] of one's prolonged existence. Surprisal-minimizing systems are self-evidencing systems (Hohwy 2016): systems that strive to bring about the evidence favoring the hypothesis that they exist, thereby prolonging their existence.

In this context, models should be understood as the set of states enclosed by a Markov Blanket (e.g. Hohwy 2017; Friston *et al*. 2020).[6] Formally speaking, a Markov Blanket is the set of nodes that, within a graph, makes the target node (or set of nodes) conditionally independent from any other node of the same graph (Pearl 1988: 97). Within the PP literature, however, they are used to determine the functional boundary that separates an organism from its environment and allows the organism to interface with it (e.g. Friston 2013).[7]

Markov Blankets can be decomposed into *active* and *sensory* states. Active states are influenced by internal and sensory states, and influence the environment. Sensory states are influenced by active and environmental states, and influence internal states (Friston *et al*. 2020).[8]

Markov Blankets are multiple and nested. Every biological system can be decomposed in sub-systems and might partake in larger systems, and each of those systems will have its own Markov Blanket: Markov Blankets can thus be individuated at every possible level of

[5] We are purposefully vague on how "models" should be understood. On internalist and representationalist readings of PP, models should be understood as *structural representations* encoded in the agent's brain (Kiefer and Hohwy 2018; 2019). Conversely, according to enactive accounts, models should be understood as "entailed" by the embodied activity of an organism (Ramstead *et al*. 2020; see also Friston 2012a: 2111 for a fromal definition of entailment). The only relevant point for our argument is that models are *identified through* their Markov Blankets.

[6] Bluntly put, the states enclosed by a Markov Blanket constitute a model of the "external" states because they can be mathematically described in two equivalent ways: either by their intrinsic dynamics, or as encoding a family of probability distributions over external states. See (Parr *et al* 2019; Friston *et al*. 2020: 9-12).

[7] Notice that this "transmutation" of Markov blankets from formal properties of graphs to functional boundaries of living and cognitive systems is far from metaphysically innocent (see Bruineberg *et al*. 2020).

[8] In (semi) formal terms, active states are the children of the blanketed node, whereas sensory states are the parents of the blanketed node. Whether co-parents of the blanketed node should be interpreted as active or sensory states is unclear (see Bruineberg *et al*. 2020).

description (e.g. Kirchhoff *et al*. 2018)

The free energy principle can be related to PP noticing that free energy can be equated with *prediction error* (e.g. Friston 2005; 2009; Buckely *et al*. 2017), a well-known neural signal posited by predictive coding accounts of neural functioning (e.g. Mumford 1992; Rao and Ballard 1999; Spratling 2017).

On the view PP offers, the mammalian central nervous system is a generative model which recapitulates how sensory signals are produced by their environmental causes.[9] Such a model is *generative*, as it can endogenously generate, in a top-down manner, the sensory signals it expects to encounter. For instance, when the agent looks at a carrot, the model will predict orange sensory signals in the visual modality.

The predicted sensory signals are then contrasted with the signal actually received, quantifying the prediction error: a signal which traverses the cortex "from the bottom-up" and that brains are tasked with minimizing. It can be minimized in several ways. First, brains can revise their predictions, accommodating the incoming signal. This corresponds to perceptual recognition (Hohwy 2013: Ch. 1 and 2; Clark 2016: Ch, 1), which is associated with a decrease of D_{KL} (Wiese and Metzinger 2017). Secondly, brains can minimize prediction error through attention, decreasing the impact on neural processing of prediction errors by lowering the signal precision (Hohwy 2013: Ch. 3; Clark 2016: Ch. 3). Heuristically, this is equivalent to ignoring noisy and unreliable prediction errors. Lastly, the brain can minimize prediction error by changing the incoming signals through movement (Hohwy 2013: Ch.4; Clark 2016 Ch.4). If one's brain predicts the sensory signals related to carrots, one good way to make these predictions come true is by looking for a carrot. This is *active inference*, and minimizes surprisal *directly*, as it directly changes the states of the organism.[10]

[9] Notice that, on the view PP offers, an animal's body *is* an environmental cause, as it is an extra-neural producer of sensory signals (e.g. limb position, heartbeat rate, state of the skin, etc.)

[10] Notice: the fact that organisms cannot evaluate surprisal does not entail they cannot minimize it. Compare: we cannot evaluate the number of hair on our heads, but we can surely change that number by pulling some hair off.

Crucially, active inference is performed by predicting, and then canceling out, the sensory consequences of one's movement (Friston 2011; Namikawa *et al*. 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*; and then cancel out the prediction error generated by those predictions. But the sensory outcomes of *m* are just the sensory states the agent would encounter, were *m* performed. Hence, 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.[11]

Notice, however, that, on the view PP offers, an agent's mastery of sensorimotor contingencies is a purely *neural* affair (Clark 2012; Seth 2014). And, in fact, contrary to sensorimotor enactivism, PP offers an indirect view of perception, in which perceptual targets are *inferred* through prediction error minimization (Frith 2007; Hohwy 2013; Wiese 2018; Drayson 2018). Thus, PP appears to adhere to a strong form of vehicle internalism: prediction error minimization is an entirely neural affair, taking place within the Markov Blanket constituted by the primary sensorimotor cortices (Hohwy 2016; 2017).

Importantly, however, PP is *not* a theory of consciousness (Seth and Hohwy forthcoming), even if it seems able to account for at least some structural aspects of consciousness (Hohwy 2013; Clark 2016: Ch. 7; Seth and Tsakiris 2018). 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).

[11] In non PP models, the sensory consequences of one's actions are computed by *forward models*: special purpose generative models responsible for reafference cancellation (Pickering and Clark 2014). Importantly, forward models have provided one of the first computational implementation of an agent's sensorimotor mastery (see Maye and Engel 2013)

We concede the point for the sake of argument. Importantly, on Kirchhoff and Kiverstein's own view, only *temporally thick* generative models qualify as consciousness supporting (Kirchhoff and Kiverstein 2019a: 106-108, see also Hobson and Friston 2014). For our purposes here, it is sufficient to say that a generative model is temporally thick *if* it periodically revisits the same set of states (e.g. if it enables one to celebrate one's birthday on an annual basis, see Friston 2018: 5-6).

But what are temporally thick generative models *made of*? What are *the material bits* (i.e. the vehicles) constituting them? The PP account seems to suggest that these vehicles are squarely located within brains. Kirchhoff and Kiverstein disagree. Let us see why.

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

"DEUTS" stands for "Dynamical Entanglement and Unique Temporal Signature". According to Clark (2009; 2013), it is the strongest argument for CVE, and Kirchhoff and Kiverstein (2019a: 36) concur. As we understand it, DEUTS is 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 - Dynamical Entanglement

Cognitive processing woves agent and environment together in a single system. This idea stems from dynamical views of cognition, according to which cognitive processing is not "sandwiched" between perception and action (Hurley 2001), but rather *constituted* by sensorimotor interactions. These interactions are better explained using the formal tools of dynamical system theory (e.g. Hurley 1998; Chemero 2009; Hutto and Myin 2012) which allow to quantitatively model and predict these interactions. This explanatory methodology, however, often forces one to model agent and environment as a single (non-decomposable)

coupled system, the evolution of which accounts for the production of cognitive outputs (Lamb and Chemero 2018). The non-decomposability of the agent-environment system *causally woves* the agent into the environment, and is said to vindicate a form of *cognitive* vehicle externalism (Palermos 2014; Kiverstein 2018).[12] How can this dynamical image of (extended) cognition be related to PP?[13]

First, Kirchhoff and Kiverstein argue that prediction error minimization is a tool servicing surprisal avoidance. But surprisal can be avoided only through active inference; 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 functionally separate agent and environment, they also *enable* the coupling of the two (Kirchhoff and Kiverstein 2019a: 65-67; see also Fabry 2017).[14] 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.[15] Thus, together, active and sensory states enable internal and external states to interlock in a *two way* interaction, which is a form of coupling. Importantly, this form of coupling can be accounted for by the formal tools of dynamical system theory in terms of *generalized synchrony* (Friston 2013; Bruineberg and Rietveld 2014; Bruineberg *et al*. 2018a; Kirchhoff and Robertson 2018; Kirchhoff and Kiverstein 2019a: 108-110; 2020 *ft*. 3).

In third place, Kirchhoff and Kiverstein (2019a: 73-76; 2019b) stress that Markov

[12] Importantly, the idea that cognition extends when cognitive outputs are jointly produced by a coupled agent environment system traces back to Clark and Chalmers (1998: 8-9).
[13] Notice, importantly, that the free energy principle allows for a straightforward dynamicist treatment (Bruineberg and Rietveld 2014; Tani 2016), which can be extended to PP (Friston and Kiebel 2009).
[14] This is not a contested point in the PP literature: even the staunchest internalists concede that Markov Blankets couple agent and environment (and that this is a nomological necessity; see Hohwy 2017)
[15] 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.

Blankets are not just multiple and nested, but also malleable and plastic. To do so, they extensively rely on Clark's (2017) metamorphosis argument. Clark invites us to consider metamorphic insects, and the functional boundary (i.e. the Markov Blanket) that separates 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 skin of the caterpillar). But if this is correct, then Markov Blankets can be "moved around"; and thus the Markov Blanket that identifies our cognitive system can shift, when the appropriate conditions are met.

Consider the former 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 such a system 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. As the relevant form of coupling at play is generalized synchrony, the "wider" blanket will be a Markov Blanket encompassing smaller Markov Blankets falling into generalized synchrony. In Kirchoff and Kiverstein's (2019a: 79-81; 2019b) view, such a Blanket identifies the relevant cognitive machinery; that is, the self-evidencing model engaged in prediction error/free energy minimization. By default, such a model encompasses the entire organism body, but, as the metamorphosis argument purportedly shows, it can extend, allowing the creation "on the spot" of extended free energy/prediction error minimizing systems. On the assumption that free energy/prediction error minimization amounts to cognition[16], thus, PP allows for extended cognitive systems.

Notice that nothing, in the picture just sketched, entails CVE. Surely, Kirchhoff and

[16] One author finds the assumption debatable, *especially* when it regards free energy. But the author is nevertheless eager to concede the point for sake of discussion.

Kiverstein (2019a: 104; 2020: 2) are eager to identify prediction error/free energy minimization with cognitive *and conscious* processing.[17] But such an identification is surely disputable: PP *just isn't* a theory of consciousness (Seth and Hohwy forthcoming). Moreover, one can surely concede that whereas the Markov Blanket of the cognitive system extends in the way just seen, the Markov Blanket of the *phenomenal machinery* does not. Maybe *that* Markov Blanket always surrounds the brain only. This is just a Markov Blanket-based rendition of an important point raised by Clark (2009): vehicle externalism about *cognition* does *not* entail CVE. And the argument based on dynamical entanglement only ensures (if it works) vehicle externalism about cognition.

At this junction, reflections on the unique temporal signature of our experience kick in, supposedly showing, *contra* Clark, that the dynamical avenue to vehicle externalism about cognition *entails* CVE.

3.2 - Unique Temporal Signature

Suppose, for the sake of argument, that the machinery of consciousness is purely neural. If this supposition is correct, and if neural states are kept constant, a subject's phenomenology will be constant, *regardless of what is going on in the environment*. Thus two subjects can be neural and phenomenal duplicates *without* being environmental duplicates. This is the familiar intuition behind brains in a vat.[18] However, this intuition is wrong when it comes to consider the temporal evolution of the phenomenal states of dynamical entangled subjects. Or so Hurley (1998: Ch. 8) argues.

To see the reason for this denial, consider a simplified rendering of one of Hurley's (1998 303-314) thought experiments. On earth, subject S is in an entirely withe room. The room contains only S and a black ball at S's right. On twin earth, subject TS (i.e. S's twin) is in the

[17] And, in general, dynamicists are eager to identify cognitive processing with conscious processing (e.g. Silberstein and Chemero 2012; Kiverstein 2016).
[18] Or, has Hurley (1998; 2010) used to call it, this is our pre-theoretical "pluggability intuition".

exact same situation. 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. Importantly, S and TS are phenomenal, neural and environmental duplicate.

Now, can S and TS be phenomenal and neural duplicates *without* being environmental duplicates, as vehicle internalism about consciousness suggests? The answer is positive: it is sufficient to, say, switch the place of TS's ball from right to left and insert in TS's eyes left-to-right inverting lenses, to keep the visual input TS receives constant. In this case, it seems correct to say that TS will experience the exact same thing S experiences. So S and TS can be neural and phenomenal duplicates *without* being environmental duplicates.

But that is possible only because S and TS are not dynamically entangled with their environment (Hurley 1998: 327). Thus, suppose that S and TS try to touch the ball when they're not environmental duplicates. They will move their right arms towards the ball (which they both see at their right) and then their phenomenal (and neural) states will diverge. For S will touch the ball, whereas TS will not. So S'n neural state will be modified by the reafferent signal (whereas TS's won't), and, as a result, S will experience what it is like to touch a ball (whereas TS won't). In this case, duplication *fails*.

To allow for the experience ensuing from the dynamical entanglement of 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 experience can be duplicated in TS. So, in order for the phenomenology of a dynamically entangled agent to be duplicated, it is not sufficient that the subject and its twin are neural duplicates. They also *need* to be environmental duplicates. But what needs to be tokened in order for a mental content (in this case, a phenomenal content) to occur is the vehicle of said content. And, in the example just considered, what needs to be "tokened" for the relevant phenomenal contents to occur includes at least some environmental factors.

Hence, if a subject is dynamically entangled, environmental factors are part of the vehicles of its consciousness (Hurley 1998: 330-335). The relevant phenomenal machinery is what Hurley (1998: 2) dubbed a *dynamical singularity*: a singular structure, in the field of causal flow, characterized through time as a tangle of numerous feedback loops of varying temporal orbits. *Contra* Clark (2009), dynamical entanglement entails CVE.

As we understand them, Kirchhoff and Kiverstein (2019a: 112-115; 2020: 5-9) make essentially the same point. Importantly, however, 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. In this view, the cognitive (and phenomenal) machinery has no fixed properties. Rather, its properties are constantly transformed by the cultural practices it participates to, and these properties have to be constantly negotiated by engaging with the surrounding cultural niche (Kirchhoff and Kiverstein 2019a: Ch. 5; 2020, see also Kirchhoff 2012). Cognitive and phenomenal systems are thus enculturated, in the sense that their participation to 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 to, in a sense, assemble the cognitive/phenomenal machinery. As they write:

> "the assembly of cognitive systems is not always orchestrated by the individual agent but is sometimes distributed across a nexus of constraints, where some constraints are neural, some are bodily, and some are environmental" (Kirchhoff and Kiverstein 2019a: 16)

In more vehicular terms, the claim seems to be that cultural practices "assemble" the cognitive system insofar as they contribute to determine the *expected precision* of the incoming sensory signal. That is, they contribute to create a system's expectation for certain very precise 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). Importantly, these cultural practices shape more or less directly one's subjective experience. Kirchhoff and Kiverstein provide a variety of examples of this. One is the ability, of appropriately enculturated subjects, to see certain stars in the Ursa Major constellation as "pointer stars", and use them to find Polaris (Kirchhoff and Kiverstein 2019a: 96-97). Here, the expected precision of the incoming sensory input allows one to "parse" the relevant visual information, allowing the expert star-watcer to construct a flow of significant visual information around which subsequent actions can be organized. Or consider the way in which learning a language allows to recognize certain soundwaves as well-defined phonemes, and how such an ability impacts one's phenomenology (Roepstorff *et al*. 2010).[19] Another language related example involves the learning and mastery of a rich vocabulary of color terms, which seems to make subjects faster in discriminating hues of color (Kirchhoff and Kiverstein 2019a: 98; Thierry *et al.* 2009).

Importantly, In Kirchhoff and Kiverstein's view the culturally-leaden modification of conscious experience is not due to the acquisition of specific neural representations. Rather, it is due to the constant agent-environment interaction; and cultural practices should be seen as elements regulating the behavior of the agent-environment coupled system:

> "[...] rule, principle and standards - the patterns of cultural practice - can be thought of as a macroscopic order parameters that evolve over longer timescales" (Kirchhoff and Kiverstein 2020: 7).

Thus, if we understand Kirchhoff and Kiverstein correctly, we should not think of enculturated brains as regular brains supplied with culturally determined representations. Rather, we should think of them as nodes in a complex web of loopy causal relations, which are constantly transformed by the culturally shaped loops traversing them.

We end our exposition of Kirchhoff and Kiverstein complex position here, mainly owing

[19] Strikingly, this case is often used as a prime example of so-called "cognitive phenomenology"; that is, the phenomenal facet of cognitive (non-sensory) mental states, see (Horgan and Tienson 2002).

to space limitations. We acknowledge that our summary is incomplete[20], and that it does not convey the entire depth of Kirchhoff and Kiverstein's overall position. But we think this emaciated summary is enough, for us, to articulate our counter-arguments.

4 - Kirchhoff and Kiverstein's defense of consciousness vehicle externalism: three problems

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

Recall: CVE is the claim that the phenomenal *machinery* is not entirely located within a subject's brain. CVE makes a claim about the *vehicles* of a subject's experience: the *material carriers* of a subject's phenomenal contents. Now, in the philosophical literature vehicles are typically identified as *concrete particulars*: that is, physical, individual entities (e.g. Shea 2018: 10; Smortchkova *et al*. 2020: 2). But cultural practices (e.g. the practices of writing, performing human sacrifices, or playing football) do not seem to be concrete particulars. They do not look like vehicles. As such, it is very doubtful that they qualify as external *vehicles* of a subject's consciousness.[21]

Perhaps focusing on a concrete case might clarify *how* cultural practices function as external vehicles of subjective experience. Thus, consider the experiment performed by Thierry *et al*. (2009), which Kirchhoff and Kiverstein (2019a: 98-99) propose as evidence of the fact that cultural practices support CVE.

The study (Thierry *et al*. 2009) involved two groups of participants which spoke two different native languages; namely English and Greek. Participants of both groups performed

[20] 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.

[21] Of course, cultural artifacts used for and produced by cultural practices (e.g. written pages, sacrificial knives, soccer-balls) are concrete particulars, and *can* be external vehicles of a subject's mental machinery. This claim is as old as the "extended mind" itself. So it seems to us that, if Kirchhoff and Kiverstein are proposing a new, radical, "third wave" form of vehicle externalism they cannot be claiming *just that*.

the same taks: participants were required to perform an odball stimulus discrimination. More in detail, 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 neuronal activity of each participant was captured through an EEG cap.

Here's the crucial bit of the experiment: albeit participants were required to discriminate *shapes*, stimuli could also vary in *color*. In total, for color were used: light and dark blue and light and dark green. Notice that the variation in color is *entirely task irrelevant*: if, say, a yellow stimulus suddenly appeared, participants were *not* expected to press the button (unless the stimulus was a circle).[22] 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 dark or light), only English uses a single word for blue. Greek uses two: *galazio* (dark blue) and *ble* (light 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 their data support the claim that color terminology can influence early visual processing.

We must confess that we simply do not see *how* this experiment is supposed to bolster CVE. Let us start with consciousness. The experiment does *not* establish that Greek and English native speakers experience color differently. As The experimenters write, their 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

[22] The yellow stimulus example is *just an example* for the sake of clarity. No yellow stimulus was actually used.

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.

Let us now move to vehicle externalism. What are the relevant external vehicles that should "extend the mind" here? 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 take a great care in describing the "external vehicles" involved, and what sort of effects they might have on cognizing. Thierry and colleagues do nothing of that sort. And they do so *rightfully*, as their experimental procedure involves *no* external vehicle.

Moreover, as Thierry and colleagues describe their experimental set-up, the subjects were not dynamically entangled to anything. Participants were only required to press a button when the oddball was detected. 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 (Clark 1997; 2008: 15-29; 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).[23]

Noticing that a subject's mastery of cultural practices has to be constantly maintained through repeated cycles of interaction with the environment (Hurley 2010: 142-143; Kirchhoff and Kiverstein 2019a; 2020: 6) does not, in our opinion, alter the dialectical

[23] Roughly the same line of reasoning, it seems to us, holds for the other examples (seeing certain stars as "pointing stars" and phoneme recognition) Kirchhoff and Kiverstein present in favor of the view that cultural practices support CVE. We here focus on the experiment by Thierry and colleagues as their experimental procedure is clearly described in a peer-reviewed journal.

situation. To see why, consider the following analogy. An athlete's muscular tone *must* be constantly maintained through repeated cycles of interaction with a culturally shaped environment (namely, a gym). But the athlete's muscular tone is entirely "internal" to the athlete. The fact that the "muscular machinery" has to constantly be finessed through environmental interactions does not entail that the "muscular machinery" is partially constituted by environmental stuff. There seems to be no external vehicle of the athlete's muscular tone. A similar conclusion, it seems to us, holds for the athlete's (or anyone else's) phenomenal machinery too.

Crucially, this vehicle internalist conclusion strikes us as being entirely consistent with the *letter* of Kirchhoff and Kiverstein's overall position. After all, they claim that cultural practices play a role in determining the *expected precision* of the various input streams (Kirchhoff and Kiverstein 2019a: Ch. 5). But, at least on the account PP offers, precision estimation is a purely neural affair, which has to do with the sharpening of neural representation and the synchronization of neuronal populations (Friston 2012b). Of course, it might be the case that precision estimation is *not* a purely neural affair, and that the machinery of precision estimation is not purely neural. But if this is the case, then Kirchhoff and Kiverstein *owe us an account* of extraneural precision estimation. As far as we can see, no such account is even sketched in their publications.

Notice that our vehicle internalist conclusion does not force us to the implausible claim that cultural practices *have no effect whatsoever* on a subject's consciousness. The case of phonetic recognition (i.e. how our perception of linguistic stimuli changes based on the languages we know) is one clear example of the effects of culture on our phenomenology (see also Lupyan *et al*. 2020 for a recent review). Indeed, the theoretical apparatus of PP seems almost ideally suited to account for such effects (see Clark 2016: Ch. 2; Hohwy 2017b). 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. As a nice example, consider the (now widely known) "white christmas" experiment by Merckelbach and van de Ven (2001). In this experimental setup, the experimenters made their subjects (a number of undergraduate students) listen to a short audio track containing *only* random noises. Crucially, however, before the stimulus was presented, the experimenters informed the subjects that a recording of "White Christmas" was "buried under" the noise. About one third of the subjects reported *actually hearing* the song - even if *no* song was actually present. This case nicely illustrates how a mastery of sociocultural practices (in this case, language) can affect a subject's phenomenology.

Notice, however, that these effects seem to operate on the *content* of a subject's phenomenology. In Merckelbach and van de Ven's experiment, for instance, the fact that the subjects expected to hear "White Christmas" changed *what the subject perceived*, not the machinery by means of which the subjects perceived.[24] A similar line of thought seems to hold for all the examples proposed by Kirchhoff and Kiverstein (i.e. seeing certain stars as "pointer stars", hearing phonemes) and the many cases discussed in (Lupyan *et al*. 2020). In brief, it seems that cultural practices can, at least sometimes, modify the *contents* of one's subjective experience. So, maybe Kirchhoff and Kiverstein's emphasis on cultural practices can vindicate (or motivate) some form of externalism about phenomenal *content*. However, externalism about phenomenal content surely does neither entail or support 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 content is determined by extraneural factors does not, in and by itself, entail that the

[24] Of course, there is a sense in which the machinery by means of which the subjects perceived changed when subjects expected to hear "White Christmas" (for instance, by generating some pattern of activity corresponding to the expected inputs). But, at least as Marckelbach and van de Ven describe their experiment, all the relevant changes seem to happen inside the subject's brain. So these changes do not seem to lend any support to CVE.

vehicle of said content is extraneural. A teleosemanticist, for instance, holds that content is partially determined by *evolutionary functions* (e.g. Millikan 1984); that is, by what a given type or device or item was selected for on an evolutionary timescale. However, this view clearly does not imply that *the vehicles* of an agent's mental representation include the agent's evolutionary ancestors (on the pain of absurdity).

Let us summarize this sub-section. We have claimed that Kirchhoff and Kiverstein's emphasis on cultural practices does not support CVE. To do so, we have scrutinized one paradigmatic case they propose, noticing that, simply put, no *external vehicle* was involved in such a case. We also noticed that Kirchhoff and Kiverstein's position on cultural practices, at least based on how they phrased it thus far, is compatible with vehicle internalism; and that cultural practices seem to impact the *contents* of subjective experience, rather than its vehicles.

We close this section with a piece of advice. Kirchhoff and Kiverstein (2019a; 2020) seems willing to propose a new account of constitution, able to "factor in" the diachronic role of cognitive practices in "extending" the conscious mind.[25] We must confess that such an account of constitution is still mysterious (at least, to us): we have not yet figured out under which conditions cultural practices are supposed to count as external vehicles of a subject's consciousness. However, if our discussion in this section is correct, it seems to us that Kirchhoff and Kiverstein *do not need* such an account. The role of cultural practices in shaping a subject's consciousness seems (at least, *prima facie*) sufficiently accounted for in terms of phenomenal contents. This strikes us as a more promising path of research - one we would gladly trot alongside Kirchhoff and Kiverstein.

[25] Importantly, Clark and Chalmers (1998) already noticed that, whereas content externalism tends to focus on extraneural factors in a subject's past, vehicle externalism focuses on the active role external resources play in the *present*. For this reason, it seems to us that Kirchhoff and Kiverstein's focus on the cultural practices a subject took part in in its past naturally suggests that cultural practices might support a form of content, rather then vehicle, externalism.

4.2 - The consciousness bloat objection

In the PP literature, the issues surrounding vehicle externalism (and thus CVE) are typically framed in terms of Markov Blankets (Hohwy 2016; Clark 2017). But there seem to be many Markov Blankets in the world.[26] Thus, the relevant question is: *which* Blanket should we pick to identify the relevant system we are interested in (in our case, the phenomenal machinery), and *why* should we pick that Blanket over any other?

As far as we can see, Kirchhoff and Kiverstein (2019a; 2019b) adhere to this Markov Blanket based framework in their defense of CVE. In fact, they propose a crisp answer to both questions. They argue that the relevant Blanket surrounding the phenomenal machinery is the wider Blanket that comprises many "smaller", synchronized Blankets in its innard. They also argue that this is the relevant Blanket because its dynamics (i.e. the way in which active and sensory states allow the coupling of internal and external states) acts as an order parameter on the "smaller" Blankets it contains, sucking them into a coordinate pattern, as mutually interacting components of a single system (Kirchhoff and Kiverstein 2019a: 80-81).

Now, leave Markov Blankets aside for a moment, and focus on vehicle externalism. A prominent objection to vehicle externalism is the so-called "cognitive bloat objection" (Rupert 2004; Sprevak 2009). The objection is basically a slippery-slope objection that points out that, given the criteria the vehicle externalist proposes to determine whether an external item qualifies as a constituent bit of the relevant mental machinery, simply *too much stuff* gets counted as a cog in the mental machinery. This leads to unpalatable consequences, such as an explosion of the mental states attributable to an agent (e. g. Ludwig 2015), and it is generally taken as a *reductio* of the relevant form of vehicle externalism.

Here, argue that Kirchhoff and Kiverstein's position form of CVE is susceptible to precisely this kind of objection. More specifically, we argue that Kirchhoff and Kiverstein's

[26] Indeed, sometimes it is claimed that the entire planet has its own Markov Blanket (Rubin *et al*. 2020).

way of identifying the relevant Markov Blanket allows to include, as cogs in a subject's phenomenal machinery, *other subjects*.[27]

To see why this is the case, consider first the (seemingly well established claim) that any two PP systems busy modelling each other rapidly fall into generalized synchrony (e.g. Palacios *et al*. 2019). Originally, this point was empirically demonstrated by Friston and Frith (2015a; 2015b) through computational simulations. The simulation itself is mathematically complex, but the idea behind it can be clearly expressed as 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 turn, B will try to predict A's moves; and, more specifically, the sensory consequences of these 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.[28]

Aside from Friston and Frith's simulations, there are other threads of evidence suggesting that interacting subjects tend to synchronize at multiple levels (see Wheatley *et al*. 2012; Coey *et al.* 2012; Tognoli *et al*. 2020 for reviews). Ecological psychologists, for instance, have long noticed that the limb movement of visually coupled subjects tend to synchronize (e.g. Schmidt *et al*. 1990; Richardson *et al*. 2005; see also Schmidt and Richardson 2008 for a

[27] Importantly, there also seems to be something odd with Kirchhoff and Kiverstein's claim that the relevant Markov Blanket identifying a subject's phenomenal machinery is, by default, placed around the entire embodied organism (Kirchhoff and Kiverstein 2019a: 80). For that Markov Blanket includes a multitude of things (e.g. kidneys, lungs, toenails) which, at least *prima facie*, we have no reason to regard as cogs of the phenomenal machinery. Notice that noting this does in no way beg the question against CVE, at least as sensorimotor enactivists conceive it. For, according to the sensorimotor enactivist CVE is established by the dynamical entanglement of subject and environment (Hurley 1998). But there seems to be just no meaningful dynamical entanglement between, say, a subject and the subject's toenails.

[28] Of course, their synchronization is not *perfect*, not even in the idealized simulations proposed by Friston and Frith. For one thing, the two systems are bound to differ on the precision assigned to these predictions see (Friston and Frith 2015a; 2015b).

review).[29] Interacting subjects also tend to synchronize their postural sway (Shockley *et al*. 2003) and some of their autonomic responses, such as their patterns of pupil dilatation (Kang and Wheatley 2017). Indeed, intrapersonal synchronization seems a very pervasive phenomenon, so much so that people sitting on rocking chairs tend to (unconsciously) synchronize the way in which they rock (Goodman *et al*. 2005; Richardson *et al*. 2007). Moreover, there is ample empirical evidence (obtained independently from the computational framework of PP) that the neural activity of interacting subjects synchronizes (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). It thus seems correct to conclude that, when two or more subjects interact, they actually tend to synchronize.

Consider now the computational simulations presented in (Palacios *et al*. 2020) and (Friston *et al*. 2015). If correct, these simulations show that when a number of 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. The prior expectations regarding perception are one clear example. For instance, it seems that humans expect natural light to illuminate objects from above and slightly on the left (e.g. Mamassian *et al*. 2002). Or, to give but another example, our prior expectations about noses being convex is so strong that it can generate the "hollow-face" illusion (i.e. seeing a *convex* face when looking at an appropriately illuminated *concave* side of a mask). Broadly speaking, it is widely recognized that, on the account of perception PP offers, perceptual systems must be properly attuned to

[29] Notably, this line of research models the limbs of the experimental subjects as interacting pendula. Curiously, the first time generalized synchrony was described (Huygens 1673), it was described as the synchronization of interacting pendula. And, just as the limbs of the participants in Shmidt and colleagues' experiments, Huygens's pendula phase-locked in antiphase.

the environmental statistics (e.g. Orlandi 2014; 2016).[30] Hence, it seems that the perceptual systems of agents inhabiting the same environment must, broadly speaking, encode the same prior expectations. The same holds true if we consider the account of *action* PP proposes, namely active inference. According to active inference, at least some, very general prior expectations motivating actions (e.g. the expectation of being well-fed) have been hardwired in the control system by natural selection, and are thus shared by a great number of agents, humans included (e.g. Friston et al 2012a: 525; 2012b; Sims 2017). Consider, further, the account of the mirror system PP offers. On this account, the mirror system stores a model of one's bodily dynamics, which can be used to predict both one's action and someone else's action, provided that the target of these predictions is sufficiently "like" the predictor (Kilner *et al*. 2007; 2011; Donnarumma *et al*. 2017). If this account is correct, agents that are sufficiently "alike" are bound to have some expectation on their bodily dynamics in common. Lastly, consider culturally established prior expectations. These practices are said to establish regimes of shared expectations among the members of a culture (Roespstorff *et al*. 2010; Constant *et al*. 2019; Kirchhoff and Kiverstein 2019a: Ch. 5; 2020). Taken together, all these threads of evidence strongly suggest that humans have at least some shared expectations. 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.

Now, if that is correct, and if it is correct (as it seems) that interacting human subjects tend to fall into generalized synchrony at multiple scales, then it surely seems correct to say that the interaction of human subjects leads to a "wider" Blanket containing multiple "smaller" Markov Blanket falling into generalized synchrony. But such a "wider" Blanket is precisely the sort of Blanket that, in Kirchhoff and Kiversteins view, identifies a subject's phenomenal

[30] More specifically, the perceptual system of each species must be attuned to the relevant regularities of the ecological niche the specie inhabits.

machinery. Thus, if Kirchhoff and Kiverstein are correct, it seems that anytime two (or more) subjects interact with each other, they all end up being counted as constituents parts of the phenomenal machinery of each other. This conclusion surely strikes many (the authors included) as mildly unpalatable.

We believe that, on intuitive grounds, such a conclusion is *so* unpalatable to constitute a *reductio* of Kirchhoff and Kiverstein's position. To us, the point seems exactly the point leveraged by Block's (1978) famous "China brain" thought experiment. No matter how cleverly arranged, *groups of subjects cannot* constitute any phenomenal machinery; and surely they cannot (partially) constitute the phenomenal machinery of one of the subject's of the group.

Now, a foreseeable objection to our claim is roughly the following: *granted*, Kirchhoff and Kiverstein's (2019a; 2019b; 2020) defense of CVE has intuitively unpalatable conclusions. But intuitive unpalatability is not a sign of falisty - there are many intuitively unpalatable, yet true, propositions. Moreover, the vehicle externalist can simply *accept* the bloat and call for a revision of the relevant underlying metaphysics (Chalmers 2019: 16). After all, vehicle externalism *is* a revisionary claim. So it should not be surprising that it has revisionary consequences.

To answer this foreseeable objection, we highlight the following: that Kirchhoff and Kiverstein (2019a: 53; 106-107) accept that only *temporally thick* models[31] can be consciousness-supporting. That is, they believe that only the vehicles making up temporally thick models can qualify as *phenomenal machineries*.

Now, 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 more a model will be temporally thick. A

[31] Recall that, in this context, "model" refers to anything enshrouded by a Markov Blanket

bacterium, for instance, might revisit a state *x* every half an our. Its model is thus significantly shallower than the model a typical human possesses, given that a typical human visit certain states only once a year (e.g. throwing a party for one's own birthday). An hypothetical creature that revisits a state only once a century will have an extremely temporally thick modal (if compared to us). This means that, in order to be temporally thick, a model must pay successive visits to the same states. A temporally thick generative model *loops* though 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 *other* human interactions do not have such a loopy structure. Indeed, some human interactions are *one-shot*. 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 Markov Blanket "surrounding" them both. But the model identified though such a Blanket will not re-visit any of its states 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.[32] However, given the way in which Kirchhoff and Kiverstein propose to identify a subject's phenomenal machinery, that

[32] Importantly, the same conclusion seems to hold true even when one considers a more regimented notion of temporal thickness, for instance in terms of a multi-layered model in which each layer predicts the incoming input at a different temporal scale (e.g. Tani 2016: 199-218; Friston *et al*. 2017). It is not at all clear, for instance, what the various layers should be in the case of the applicant-job interviewer dyad.

model is identified as a piece of phenomenal machinery. This, it seems to us, poses a dilemma to Kirchhoff and Kiverstein: either the claim that only temporally thick generative models are consciousness-supporting is incorrect and oughts to be rejected, or the way in which they identify "extended" phenomenal machineries is incorrect, and oughts to be rejected. At any rate, Kirchhoff and Kiverstein must give up *some* claim that they endorse, regardless of pre-theoretical intuitions on what could possibly qualify as phenomenal machinery.

4.3 - DEUTS entails the consciousness bloat

Recall the general structure of the DEUTS argiment. The first step is a commitment to dynamicsm: cognitive processes are often constituted by agent-environment sensorimotor interactions. The best way to explain these interactions is through the tools of dynamical system theory; but, 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; Hutto *et al*. 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 extended experience to be experienced, that agent and environment interact in a certain way; and thus that certain environmental features qualify as external vehicles of experience. As Susan Hurley magistrally wrote:

> "The subpersonal states and processes that do include all the token-explanatory factors, with nothing left out, should in principle be duplicable in a different environment. If certain subpersonal states or

> processes are not duplicable, then they do not include everything that is doing token-explanatory work. Something playing a token explanatory role has been left out; the boundaries around the token-explanatory states, *or vehicles*, should be expanded. [...] So, oversimplifying for clarity: if a vehicle, then duplicable, *and if not duplicable, then not the whole vehicle*" (Hurley 1998: 331, emphasis added).

Hence, if at least some environmental factors *need* to be duplicated in order to duplicate a phenomenal experience, then the phenomenal vehicles will be partially spread in the environment. And this, of course, just is CVE. Crucially, Hurley takes this line of thought to provide a *discriminating* way to appeal to the causal spread of cognitive processing (Hurley 1998: 330). She knew that each and 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, the chair upon which she seats, the feeble gravitational attraction the star *Altair* is exerting upon her, etc); and she knew that *not all* these causal interactions could plausibly be counted as constitutive parts of a subject's phenomenal machinery. That would be a reductio of CVE.

To make her appeal to the causal spreadness of cognitive processing discriminating (that is, sensitive to the difference between mere environmental causes and genuine external vehicles), she 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 be the case in order for a given piece of phenomenology to be the case is the vehicle of that piece of phenomenology. And, in the case of dynamically entangled subjects, what needs to be the case in order for certain bits of phenomenology to be experienced includes environmental (broadly speaking, extraneural) features. Hence, CVE is correct.

Here, we wish to argue that Hurley's line of reasoning is not *discriminating enough*. In particular, we wish to claim that DEUTS forces one to identify *at least some* background

causal factors as genuine constituents of a subject's phenomenal machinery. To do so, we propose a thought experiment in the style of those discussed by Hurley (1998) herself, and briefly presented in this paper in section 3.

So, let us consider a subject S and its duplicate TS on twin earth. Let us suppose that they are in entirely white rooms; and that, alongside the subjects, these rooms only contain a black ball, located at the right of both subjects. The only difference between S's condition and TS's condition is the amount of oxygen[33] present in the room: whereas S is in a normally oxygenated room, TS isn't. In fact, there is no oxygen in TS's room. Lastly, let both S and TS be dynamically entangled with their balls: they both want to juggle it in the air.

It seems clear that, given the proposed setup, the phenomenology of S and TS 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. Hence, it seems we cannot duplicate S's experience in TS, if the environment does not contain enough oxygen. However, as Hurley wrote: "if not duplicable, then not the whole vehicle". We should thus look for a further factor, the occurrence of which is needed for the relevant phenomenology to be experienced. But given the simplicity of the thought experiment proposes, that factor can only be the oxygen. And, in fact, were the oxygen present in TS's room, TS too would feel what it is like to play with the ball. So, when oxygen is included, S's experience can be duplicated. But if the DEUTS argument is correct, this implies that the oxygen is part of S's phenomenal machinery, which surely is incorrect.

A sensorimotor enactivist willing to defend DEUTS might, at this point, just try and bite the bullet. Maybe our thought experiment is the only thought experiment that yields such an undesirable consequence. The problem with this line of reasoning, we believe, is that with enough creativity *almost everything* ends up being counted as a constituent of a subject's phenomenal machinery. Thus suppose that S and TS are in their white (equally oxygenated)

[33] 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.

rooms, playing with the balls. Suppose now S's ball was painted with a special paint, which emanates the smell of flowers when heated, whereas TS's ball was not. Since both S and TS are furiously playing with their respective balls, their balls heat up, making S sense the smell of flowers. TS, however, feels no such smell. So S's experience is, again, anot duplicated. Paint TS's ball with the same paint, and the experience will be duplicated. So, if DEUTS is right, that paint is part of S's phenomenal machinery. Examples of this sort seem *dangerously* easy to come by. Bluntly put, the problem is this: the DEUTS argument seems to allow each and every factor which (more or less directly) partially contributes in determining a subject's phenomenology to count as a constituent part of the subject phenomenal machinery. For this reason, we claim that the structure of the DEUTS argument fundamentally entails a consciousness bloat. For this reason, we also believe, *pace* Clark (2009) and Kirchhoff and Kiverstein (2019a; 2020), that DEUTS is not the strongest argument in favor of CVE.

But perhaps DEUTS might be supplemented by some further criterion, which might enable one to tell apart genuine constituents from mere causal factors, yielding a better defense of CVE?

We will soon consider (and attack) the two additional criteria proposed by Kirchhoff and Kiverstein (2019b; 2020). But before doing so, we wish to highlight the following point. Suppose that DEUTS can be supplemented by a further criterion *C*, which enables a proper defense of CVE by telling apart the constitutive bits of a subject's phenomenal machinery from mere causal factors impinging upon a subject's consciousness. Now, if our analysis of DEUTS is on the right track, it seems correct to conclude that what is doing "all the hard work" in establishing CVE is *C*, rather than DEUTS. That is: if DEUTS does not discriminate between constituent and merely causal factors but *C* does, it is *C* what determines whether a candidate external vehicle of consciousness *really* counts as a vehicle of consciousness. In such a scenario, it seems to us, the DEUTS argument *per se* would be doing no useful work

in establishing and defending the truth of CVE.

That being said, what are the additional criteria that might save, if not DEUTS, at least CVE? To our knowledge, Kirchhoff and Kiverstein propose two further criteria aimed at avoiding bloat-style objections. Curiously, they never propose them *in conjunction*, so we will deal with them separately.

The first criterion is proposed in Kirchhoff and Kiverstein (2019b: 16-18).[34] The idea seems to be the following: an external candidate vehicle 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).

This criterion has some initial plausibility. It surely prevents many *candidate* vehicles from being counted as *actual* vehicles of a subject's phenomenal states. Consider, for instance, the objection we raised in section 4.2. There, we claimed that the Markov Blanket surrounding two interacting subjects seems to qualify, according to the criteria proposed by Kirchhoff and Kiverstein, as a subject's phenomenal machinery. We also highlighted that such a conclusion does not sit nicely with Kirchhoff and Kiverstein's claim that only temporally thick models can plausibly qualify as consciousness supporting. The problem, as we phrased it above, was the following: many interactions among subjects happen *only once*, so the model the two subjects jointly constitute will never revisit some state in the future, and hence it cannot have any relevant temporal thickness. It should be clear how the additional constraint required in Kirchhoff and Kiverstein (2019b) blocks our objection: since occasional interactions cannot, bing occasional, keep surprised at minimum over time, the

[34] To be fair, in that paper Kirchhoff and Kiverstein *do not* detal with CVE directly, but only with *cognitive* vehicle externalism. But, if DEUTS is correct, cognitive vehicle externalism *entails* CVE, so the relevant additional criteria a candidate vehicle must satisfy in order to be properly counted as an external vehicle of cognition seem to at least partially determine whether a candidate vehicle *of consciousness* really counts as an external vehicle of consciousness.

putative external vehicles of consciousness involved in these interactions (i.e. other subjects) do not qualify as vehicles. Hence, the consciousness bloat is avoided. We see, however, two problems with this criterion.

First, it clashes with the "metamorphosis argument", whose conclusion Kirchhoff and Kiverstein (2019a: 73-76; 2019b) wish to endorse. Even more generally, this criterion clashes with the claim that the boundaries of the mind are flexible and always open for renegotiation (Kirchhoff and Kiverstein 2019a: 16). The reason is fairly simple: if the only candidate vehicles that really qualify as external vehicles are the ones that keep surprisal at minimum over time[35], then there seems to be just no way to re-negotiate the boundaries of the mind "on the fly", so as to include some *temporarily* relevant external prop in the mental machinery.

The point can be made more precise in the following way. Consider some paradigmatic cases of extended cognition, such as the pressing of the "rotate" button while playing the video game Tetris (Clark and Chalmers 1998) or the usage of pen and paper to do math (Wilson 1994). Now, it seems a plain fact that we do not *always* (or even typically) engage with the relevant external props mentioned in those examples when solving cognitive tasks. Sometimes (maybe *most* of the time) we do our math "in our head". And, unless one is a *compulsive* player of Tetris, it seems very unlikely that the "rotate" button plays a role in surprisal minimization overtime. More generally, proponents of vehicle externalism tend to stress that our cognitive machinery is opportunistic, recruiting appropriate external vehicles "on the spot", as the need arises (Clark 2008). But the criterion proposed by Kirchhoff and Kiverstein (2019b) seems to prevent those external resources from counting as genuine external vehicles.

Secondly (and, perhaps, more problematically), the criterion proposed by Kirchhoff and Kiverstein (2019b) does *not* avoid the conscious (or cognitive) bloat. To see why, consider

[35] Or, as Hohwy (2016) would put it, "on average and in the long run"

interoceptive active inference. If the view of the brain PP offers is correct, brains are prediction engines: systems 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 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 interoceptive signals.[36] 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 case of interoceptive active inference occurs when humans predict their bodily temperature being around 36.6°. When prediction error relative to this prediction ensues, it can be minimized through active inference in various ways. For example, one might change clothes to adjust with the external temperature. Changing clothes will minimize the relevant prediction error. And clothes play this role *overtime*: we humans are more often than not dressed, precisely because we need to avoid the surprisaling sensory states that ensue when our bodily temperature significantly diverges from 36.6°. 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 cognitive bloat objection is not avoided.

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

Let us now examine 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[37] 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).[38]

We must confess that we do not see how this simple test can help Kirchhoff and Kiverstein's cause, as the D_{KL} is typically associated with perceptual inference(Wiese and Metzinger 2017; Wiese 2018; Bruineberg *et al*. 2018; Kiefer and Hohwy 2018; 2019). If a subject's correctly determines the external cause of the incoming sensory signals, the subject's D_{KL} will lower. Conversely, if a subject mis-infers the cause of the incoming signal, the subject's D_{KL} will rise, as the subject's "best guess" (technically, the subject's posterior distribution) is not a good approximate of the correct posterior distribution. Now, if this is correct, it is obvious to conclude that any intervention on an external cause of the sensory

[37] Technically speaking, they are encoded in the agent's recognition density.

[38] 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 impact that machinery. Importantly, however, Kaplan's criterion requires at least two counterfactual interventions. In his view, for a putative external vehicle to qualify as a genuine constituent of a subject's mental machinery, it must be the case that "bottom-up" interventions on the putative vehicle end up impacting the relevant functioning of the machinery; *and* that "top-down" interventions on the relevant functioning of the machinery end up impacting the putative vehicle. 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.

signal will change a subject's D_{KL}, and presumably the subject's phenomenology.

If this is the case, then the criterion proposed by Kirchhoff and Kiverstein (2020) is even *less* discriminating than the original DEUTS argument. For, according to the DEUTS argument, something qualifies as an external vehicle of consciousness if a subject is dynamically coupled to it. According to the DEUTS argument, for something to qualify as an external vehicle of consciousness, it must be coupled to the subject through a dense loop of continuous reciprocal causation (see Hurley 1998; Palermos 2014; see also Clark 1997).

But surely such a loop is *not* involved in all instances of perception. One can just sense a perceptual object without *affecting it in any way*. When this happens, surely the perceptual object is affecting the perceiver's sensorium (and neural activity) through some sort of causal connection. But there is no causal arrow starting from the subject and landing on the object. When we stare at the sunset, we *do not* exert any causal power upon the sun.[39]

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

5 - Conclusion

In this essay, we have examined some aspects of Kirchhoff and Kiverstein's DEUTS-based marriage of PP and CVE. We have extensively argued 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, it seems correct to conclude that, contrary to a popular opinion (e.g. Clark 2009), DEUTS is not the best argument in favor of CVE within the sensorimotor enactivists' arsenal.

Does this imply that CVE is simply *false*? No, it does not. There are other arguments in

[39] Notice that this observation does not imply that no movement is involved in "passive" instances of perception. To continue with the example above, when we gaze at the sunset, we surely perform saccadic eye movements. But these movements do not affect the sun in any way.

favor of CVE (Vold 2015; Farkas 2019) which might succeed where DEUTS fails. But, thus far, very little attention has been paid to those arguments. Perhaps due to Clark's (2009; 2013) *almost favorable* judgment, DEUTS has always been the focus of the debate on CVE (e.g. Ward 2012; Pepper 2014). We thus suggest that time is ripe to put DEUTS in retirement, and find some new argument in favor of CVE. If the arguments we presented here are on the right track, vehicle externalist have *a lot* of work to do; and the move, now, is theirs.

References

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

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.

Block, N. (1978). Troubles with functionalism. In A. I. Goodman (Ed.), (1993), *Readings in Philosophy and Cognitive Science*, (pp. 231-253). Cambridge, MA.: The MIT Press.

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., *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, D. (2019). Extended cognition and extended consciousness. In M. Colombo, E. Irvine, M. Stapleton (Eds.), *Andy Clark and His Critics* (pp. 9-20). New York: Oxford University Press.

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

Clark, A. (1997). The dynamical challenge. *Cognitive Science*, 21(4), 461-481.

Clark, A. (2008). *Supersizing the Mind*. New York: Oxford University 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. (2010). Memento's revenge. In R. Menary (ed.), *The Extended Mind*, (pp. 43-66). Cambridge, MA.: The MIT Press.

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

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

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

Drayson, Z. (2018). Direct perception and the predictive mind. *Philosophical Studies*, 175(12), 3145-3164.

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

Farkas, K. (2019). Extended mental features. In M.Colombo, E. Irvine, M. Stapleton (Eds.), *Andy Clark and His Critics*, (pp. 44-55). New York: Oxford University Press.

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., *et al.* (2012a). Active inference and agency: optimal control without cost functions. *Biological Cybernetics*, 106(8-9), 523-541.

Friston, K., *et al*. (2012b). Free-energy minimization and the dark room problem. *Frontiers in Psychology*, 3: 130.

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

Friston, K., & Kiebel, S. (2009). Predictive coding under the free energy principle. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364(1512), 1211-1221.

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

Frith, C. (2007). *Making up the Mind*. Oxford: Blackwell.

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.

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. (2017a). 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. (2017b). Priors in perception: top-down modulation Bayesian perceptual learning rate, and prediction error minimization. *Consciousness and Cognition*, *47*, 75-85.

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

Horgan, T., & Kriegel, U. (2008). Phenomenal intentionality meets the extended mind. *The Monist*, *91*(2), 347-373.

Horgan, T.; & Tienson, J. (2002). The intentionality of phenomenology and the phenomenology of intentionality. In D. J. Chalmers (Ed.) *Philosophy of Mind: Classical and Contemporary Readings* (pp. 520-533). New York: Oxford University Press.

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. (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.

Hutto, D., & Myin, E. (2012). *Radicalizing Enactivism*. Cambridge, MA.: The MIT Press.

Hutto, D., Kirchhoff, M. D, & Myin, E. (2014). Extensive enactivism: why keep it all in?. *Frontiers in Human Neuroscience*, 8: 706.

Huygens, C. (1673). *Horologium Oscillatorium*. France: Parisiis.

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.

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

Kiefer, A., & Hohwy, J. (2019). Representation in the prediction error minimization framework. In S. Robins, J. Symons, P. Calvo (Eds.), *The Routledge Companion to Philosophy of Psychology* (2nd Ed.) (pp. 384-410). New York: Routledge.

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. (2012). Extended cognition and fixed properties: step towards a third wave version of extended cognition. *Phenomenology and the Cognitive Sciences*, 11(2), 287-302.

Kirchhoff, M. D. (2015). Species of realization and the free energy principle. *Australasian Journal of Philosophy*, 93(4), 706-723.

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., & Robertson I. (2018). Enactivism and predictive processing: a non representational view. *Philosophical Explorations*, *21*(2), 264-281.

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. (2016). The interdependence of embodied cognition and consciousness. *Journal of Consciousness Studies*, 23(5-6), 105-137.

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.

Kelso, S. (1995). *Dynamic Patterns. The Self-Organization of Brain and Behavior*. Cambridge, MA.: The MIT Press.

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.

Ludwig, D. (2015). Extended cognition and the explosion of knowledge. *Philosophical Psychology*, *28*(3), 355-368.

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.

Merckelbach, H., & van de Ven, V. (2001). Another White Christmas: fantasy proneness and "hallucinatory experiences" in undergraduate students. *Journal of Behavior Therapy and Experimental Psychiatry*, 32, 137-144.

Millikan, R. G. (1984). *Language, Thought and Other Biological Categories*. Cambrdige, MA.: The MIT Press.

Mumford, D. (1992). On the computational architecture of the neocortex. *Biological Cybernetics*, 66(3), 241-251.

Namikawa, J., *et al.* (2011). A neurodynamic account of spontaneous behavior. *PLoS Comput Biol*. 7(10): e1002221.

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

Noë, A. (2006). Experience without the head. In T. S. Gendler, J. Hawthorne (Eds.), *Perceptual Experience*, (pp. 411-434). New York: Oxford University 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. (2012). How to build a robot that is conscious and feels. *Minds and Machines*, 22(2), 117-136.

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.

Orlandi, N. (2016). Bayesian perception is ecological perception. *Philosophical Topics*, *44*(2), 327-352.

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.

Parr, T., *et al*. (2019). Markov blankets, information geometry and stochastic thermodynamics. *Philosophical Transactions of the Royal Society A*, *378*: 20190519.

Pearl, J. (1988). *Probabilistic Reasoning in Intelligent Systems*. San Francisco: Morgan Kauffman.

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.

Pickering, M. J., & Clark, A. (2014). Getting ahead: forward models and their place in cognitive architecture. *Trends in Cognitive Sciences*, 18(9), 451-456.

Ramstead, M., J, D., *et al*. (2020). A tale of two densities: active inference is 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

Rubin, S., *et al*. (2020). Future climates: Markov Blankets and active inference in the biosphere. *Journal of the Royal Society Interface*, *17*:20200503.

Rupert, R. (2004). Challenges to the hypothesis of extended cognition. *The Journal of Philosophy*, 101(8), 389-402.

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.

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.

Sims, A. (2017). The problems with prediction: the dark room problem and the scope dispute. In T. Metzinger, W. Wiese (Eds.), *Philosophy and Predictive Processing*: 23. Frankfurt am Main: The MIND Group. https://doi.org/10.15502/9783958573246

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. (2015). The cybernetic bayesian brain. In T. Metzinger, J. Windt (Eds.), *Open MIND*, 35(T). Frankfurt am Main, the MIND Group. https://doi.org/10.15502/9783958570108.

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. (forthcoming). Predictive processing as a systematic basis for identifying the neural correlates of consciousness. Preprint. Retrieved at: https://psyarxiv.com/nd82g/. Last accessed 28/09/2020.

Seth, A. K., & Tsakiris, M. (2018). Being a beast machine: the somatic bases of selfhood. *Trends in Cognitive Sciences*, 22(11), 969-981.

Shea, N. (2018). *Representation in Cognitive Science*. New York: Oxford University Press.

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.

Silberstein, M., & Chemero, A. (2012). Complexity and extended phenomenological-cognitive systems. *Topics in Cognitive Science*, 4(1), 35-50.

Smortchkova, J., Dolega, K., & Schlicht, T. (Eds) (2020), *What are Mental Representations?*. New York: Oxford University Press.

Spratling, M. W. (2017). A review of predictive coding algorithms. *Brain and Cognition*, *112*, 92-97.

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.

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.

Ward, D. (2012). Enjoying the spread: conscious externalism reconsidered. *Mind*, 121(483), 731-751.

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*. Cambrdige, 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.

Wilson, R. A. (1994). Wide computationalism. *Mind*, 103(411), 351-372.