Phenomenal transparency, Cognitive extension, and Predictive processing.

Marco Facchin [corresponding author]

Istituto Universitario di Studi Superiori IUSS Pavia

Department of Human and Life Sciences; Pavia, Italy

Palazzo del Broletto, Piazza della Vittoria n. 15, 27100, Pavia

marco.facchin@iusspavia.it

https://orcid.org/0000-0001-5753-9873

Abstract:

I discuss Clark's predictive processing/extended mind hybrid, diagnosing a problem: Clark's hybrid suggests that, when we use them, we pay attention to mind-extending external resources. This clashes with a commonly accepted necessary condition of cognitive extension; namely, that mind-extending resources must be phenomenally transparent when used. I then propose a solution to this problem claiming that the phenomenal transparency condition should be rejected. To do so, I put forth a parity argument to the effect that phenomenal transparency cannot be a necessary condition on cognitive extension: roughly, since internal cognitive resources can fail to be transparent when used, by parity, external resources can fail to be phenomenally transparent too. Further, I argue that phenomenal transparency is not even a reliable indicator of cognitive extension; hence its absence should not be considered a problem for Clark's extended mind-predictive processing hybrid. Lastly, I consider and allay a number of worries my proposal might raise, and conclude the paper.

Keywords: Extended Mind, Predictive processing, Phenomenal transparency, Attention, Active inference, Recruitment puzzle.

The paper is formatted in comics sans because it is apparently more readable by various groups of people

Declarations:

1 - Introduction

The extended mind thesis claims that a subject's cognitive system is sometimes partially constituted by the bodily actions and/or the environmental scaffolds the subject relies on to solve cognitive tasks (Clark and Chalmers 1998). If this thesis is correct, the physical cogs, or vehicles[1], constituting the cognitive machinery are not all "in the head": bodily actions and environmental scaffolds are, at least sometimes, vehicles too.

Andy Clark suggests that the extended mind thesis should be merged with the predictive processing (PP) neurocomputational account of cognition. He argues that PP *complements* the extended mind thesis, providing an extension-friendly *positive* account of neural processing (Clark 2016; 2017a, b, Nave 2020). More in detail, Clark claims that PP solves the so-called "recruitment puzzle" for the extended mind; that is, the challenge of specifying *how and why* certain external resources are included in an extended problem-solving whole *without* depicting the brain as a "central executive"; i.e. a central and *exclusive* locus of cognitive control (Clark 2008: 137 *ff*).

---

[1] Here, I will use "vehicle", "resource" and "constituent (of our cognitive system/mind)" roughly as synonyms. They all refer to the *physical bits and pieces* making up an agent's cognitive machinery. Thus, the usage of "vehicle" bears no commitment to representationalism (see Hurley 1998; Menary 2007 for a similar usage). This is important to notice, for it is currently unclear whether predictive processing posits representation. See (Williams 2017; Kiefer and Hohwy 2018) for representationalist readings, and (Downey 2018; Facchin 2021a) for anti-representationalist ones.

Here I diagnose and solve a problem in Clark's PP/extended mind hybrid. The problem is that, if a widely accepted necessary condition on cognitive extension is accepted, the Clark's hybrid is inconsistent. The solution I will propose is to reject that condition, which I here dub "*phenomenal transparency condition*" (PTC). According to (PTC), mind-extending resources must be phenomenally inconspicuous, in a way I will further unpack below (§3).

I anticipate here that my arguments against (PTC) will have little to do with PP's theoretical apparatus. This will be both a boon and a curse. A curse, for (as the reviewers rightly noticed) there's a clear place, in my argument, where I "switch gears" in a potentially confusing manner (between §§3 and 4). I will do my best to dispel any potential confusion by providing an appropriate meta-commentary when the "switching of gears" happens. However, my non reliance on PP is also a boon - for, if my arguments are correct, my conclusion will be importantly independent from PP (or PP/extended mind hybrids), and will thus pose a *general* challenge to a (PTC) and connected ideas in the "4E" movement. Further, I will argue below (§ 6.4) that my proposed solution is superior to an alternative solution which heavily relies on the theoretical apparatus of PP.

Here's the outline of my plan: the next section presents Clark's PP-extended mind hybrid.[2] §3 introduces (PTC). I present my two arguments against it in §§ 4-5. §6 deals with a range of objections. §7 concludes the paper.

2 - Extended predictive minds: Clark's way

PP depicts the mind as engaged in a *single fundamental task*, that of minimizing the mismatch (or "prediction error'') between the sensory signals *expected* and the ones *actually received*. There are two ways to minimize it. One consists in revising the expectations, fitting them to the signals received. This, roughly, corresponds to perception (and learning, on higher timescales). The other consists in changing the signals received so as to encounter the predicted ones. This process (called *active inference*) is realized via a series of bodily changes tailored at altering some relevant aspect of the incoming sensory flux; but here we can simplify and consider it identical to *action* (Friston 2005; 2009; 2010; Seth and Friston 2016).

According to PP, the relevant expectations are due to the workings of a *hierarchical probabilistic generative model*: a statistical model capturing how worldly states of affairs and sensory inputs relate. Clark conceives these models as *action-oriented representations*. Generative models are

---

[2] Since the introduction is instrumental to my discussion of Clark's hybrid I will discuss PP from Clark's perspective, staying silent on alternative interpretation of (a) PP (cf. Hohwy 2013; Bruineberg 2018) (b) the relations between PP and "4E cognition" (cf. Hohwy 2016, 2017; Bruineberg et al 2018; Di Paolo et al 2021) and (c) alternative formulations of the extended mind thesis (Sutton 2010; Gallagher 20018; Kiverstein 2018) and (d) alternative ways to relate PP and the extended mind(Kirchhoff and Kiverstein 2019; Constant *et al*. 2019). I will also stay silent on so-called "Markov blankets" and the role they may play concerning the extended mind (see Menary and Gillett 2020; Kirchhoff and Kiverstein 2020; Facchin 2021b for some in-depth material on that matter).

thus depicted as "action oriented maps" (cf. Williams and Colling 2017)[3], encoding only the knowledge the agent needs in order to encounter the inputs corresponding to its pragmatic success (Clark 2015b, 2015c; 2016; see also Tschantz *et al.* 2020). As Clark says, they give agents a *pragmatic grip* on the situation agents find themselves in.

Alongside predictions and prediction errors, PP has a third major ingredient: *Expected precision.*[4] Think of it as a *weight* on prediction error based on its *expected informational quality*. If a prediction error is expected to be precise, it will be bolstered, allowing it to change subsequent neural processing, just like we would change our conduct following the negative feedback provided by a *trustworthy* source. Conversely, prediction error signals with low precision will be dampened, limiting their impact on cortical processing - just like we would ignore an untrustworthy feedback. (Feldman and Friston 2010).

Expected precision is crucial in Clark's PP/extended mind hybrid. By determining the weights on prediction errors, expected precision determines how neural regions influence each other via prediction and error signals (technically, their *effective connectivity*, see Friston 1994).

---

[3] Thus notice that, on this view, generative models are not just unstructured *action-mediating inner states* (cf Markman and Dietrich 2000a,b; Nolfi 2002). They are thus safe against prominent anti-representationalist arguments based on the simplicity and unstructuredness of action oriented/minimal representations (cf. Dreyfus 2002; Gallagher 2008; 2017 Ch.5). Curiously, one could even read some *anti-representationalists* as making this point; that is, as stressing that generative models are so structured that treating them as *mere states* is doing them a disservice (see Robertson and Kirchhoff 2019).

[4] Notice, however, that while expected precision is a single *computational* ingredient, it operates through a variety of mechanical means, ranging from the synchronization of neuronal responses to the release of neurotransmitters, see (Friston 2012; Friston *et al.* 2012a).

Clark interprets this influence as aimed at soft-assembling the coalition of neuronal resources best suited to deal with the task at hand. Thus he sees expected precision as *an inner recruiting mechanism*, busy "creating on the spot" the neuronal coalition affording the best *pragmatic grip* on the agent's current situation (Clark 2016: Ch.5).

Crucially, the deployment of such coalitions can be triggered by task-specific cues (Clark 2019: 288-290; Constant *et al.* 2021), enabling the agent to swiftly deploy intraneural informational channels allowing the effortless intake and smooth usage of environmental information (Clark 2016; 2017a,b).[5] When this happens, the agent *sense to couple with the environment* (cf Clark 2008: 15-17), in a way that closely connects PP and the extended mind thesis:

> "[...] actions that engage and exploit specific external resources will now be selected in *just the same manner as the inner coalitions of neural resources themselves*. For example, consider the case where salient high-precision information is available by the use of some bio-external device, such as a laptop or smartphone. The core routine that selects actions to reduce prediction error will now select actions that invoke the bio-external resource. *Invoking a bio-external resource, and moving our own effectors and sensors to yield high-quality task-relevant information, are here expressions of the very same underlying strategy: one that reflects our brain's best (sub-personal) estimates of where and when reliable task-relevant information is available''* (Clark 2017a: 745, emphasis added).

---

[5] It is perhaps worth pointing out explicitly that, in the PP scheme, the forward flow of environmental information is factually replaced by the flow of prediction errors, see (Clark 2015b). Indeed, prediction error just is "filtered" environmental information.

Assignments of expected precision thus do not just create task-specific coalitions of *inner* resources. They can also "weave" external resources into an agent's cognitive routine, thereby extending the agent's mind. A general solution of the "recruitment puzzle" is thus at hand: external resources are selected and integrated in extended cognitive systems *by* the expected precision mechanism *because* they are expected to yield highly precise information (cf. Friston *et al.* 2016a; 2016b).

Expected precision also plays a psychological role, determining the allocation of attention (Hohwy 2012; 2013; Clark 2016, Ch. 2). There's an intuitively clear sense in which, if all the prediction errors coming from a target *T* are systematically bolstered (at the expense of all other prediction errors) a subject's mental operations will be aimed at *T*, which will thus be the focus of the subject's mental spotlight.

Expected precision accounts for attention roughly as follows. When wordly features forcefully grab our (exogenous) attention, this happens because *large* bursts of prediction errors always have a high expected precision (Feldman and Friston 2010: 23). Attention is instead endogenously allocated when we have very clear expectations due to some underlying regularity. If, for example, the presence of some task-relevant bit of information is reliably signaled by some indicator (e.g. an arrow pointing to it), its presence will "trigger" the expectation of very precise

information incoming, making the agent *attend* certain specific incoming signals (Feldman and Friston 2010).

For my purposes here, the key point is that, according to Clark, external resources are recruited in extended minds *because* they yield highly precise prediction errors. Hence, if PP is correct, *we pay attention to them*. And this *prevents them from being cognitive extensions*, at least given a very widespread and uncontroversial necessary condition on cognitive extension.

3 - The phenomenal transparency condition

The extended mind thesis must not entail that *everything* we interact with is part of our minds.[6] That would either count as a *reductio* of the extended mind, or it would make the claim *trivial* (to claim *x* extends the mind is trivial, if everything does) (Sprevak 2009; Allern-Hermanson 2013). Hence the need for criteria separating mind extending resources from the rest of the environment.

One popular criterion is the *phenomenal transparency condition* (PTC) (see Clark 2003; 2008: 37-39; 2015a; Thompson and Stapleton 2009; Silberstein and Chemero 2011; Kiverstein and Farina 2012; Kirsh 2019; Wheeler 2005; 2019a; Piredda and di Francesco 2020):

---

[6] "Extensive enactivists" (Hutto and Myin 2013; Hutto, Kirchhoff and Myin 2014) might beg to differ: on their view, our minds is *essentially* interactive, in the sense of it *being made up* by our interactions. Yet I doubt that, if pressed, they would avoid edging their claim with some sort of limitation. If I fall down the stars I'm interacting with my environment. But I doubt extensive enactivists would like to count *the staircase* as a vehicle of my mentality!

> **Phenomenal Transparency Condition (PTC)**: a resource R partially constitutes the cognitive system of a subject S only if (among other things) R is phenomenally transparent to S.

(PTC) is a *necessary* condition. But it's not sufficient: it must operate with other conditions ("among other things"), such as, for example, "trust & glue" conditions (Clark 2010).

What does "phenomenal transparency" amount to? I propose the following characterization:

> **Phenomenal Transparency (PT)**: R is phenomenally transparent to S only if, when S uses R, R is not an intentional object S is thematically directed at; i.e. the usage of R is effortless, unconscious and automatic in a way such that S does not consciously control, monitor, pay attention to, reflect on, reason about, or think (in the broadest possible sense) about R.

(PT)[7] seems to capture phenomenal transparency as is at stake in the debate (cf. Clark 2003; 2008: 37-39; Wheeler 2019a). (PT) has 2 distinct components.[8] One is *procedural* (cf. Heersmink 2015; Grush and Springle 2019): if R is phenomenally transparent, its deployment is *effortless, unconscious and automatic*. The usage of R is a means for S to solve a problem, *not a problem S needs to solve*. When R is procedurally transparent, its deployment is as easy and effortless as the deployment of eyes to see, hands to grasp and neurons to think.[9]

---

[7] Throughout the text, I will use "(PT)" to refer to my characterization of phenomenal transparency, and use "phenomenal transparency" to indicate the property itself.

[8] More precisely: *at least* two distinct components. There may be more. But they won't play a role here (if they exist), so I don't discuss them or wonder whether they exist.

[9] Notice, however, that such a form of automaticity does not necessarily imply that a subject is an expert (in Dreyfus's 2002b; Dreyfus and Dreyfus' s 1980 sense) in using R and R-related activity. We can transparently use a pencil even if we are not expert drawers! See further §6.3 below.

The second component is broadly *phenomenological*: *S* need not be thematically intentionally directed at *R*. *R* thus must be "dropped off" *S's* consciousness, leaving her to focus on the problem she is confronting by using *R*. More precisely, *R* should not be just "dropped off" *S's* consciousness: it should also become a means for *S* to be intentionally directed to (and conscious of) other things. Wheeler (2019a) usefully explicates this saying that the transparent usage of *R* shifts the interface between subject and the environment in a way such that *R* *partially constitutes that interface*. When this happens, *R* is a means *through which* the environment is encountered in the same sense neurons are. Two paradigmatic examples may clarify the idea further.

Heidegger (2927/1961; see also Clark 2003: ch. 2) famously describes appropriately deployed tools as "ready-to-hand": they are deployed effortlessly and automatically, and the user's mind is not intentionally directed at the tool itself, but rather the task the tool enables the user to accomplish. In his view, during tool use, tools lose their phenomenal transparency only when their effortless deployment becomes impossible, perhaps due to the tool malfunctioning. When this happens, the user gets intentionally directed at the tool, that is thus "present-at-hand", as an object open to the user's conscious scrutiny. Similarly, Merleau-Ponty (1945; see also Wheeler 2019a; Kirsh 2019) contends that a blind person using a stick to perceive the environmental layout does not perceive the

stick's features (e.g. the weight, texture or temperature of the stick). Rather, the person *perceives through* the stick, just as non-blind people *see through* their retinae.

The second paradigmatic example is that of the *incorporation of tools in the body schema* (e.g. Clark 2008: 37-39; Menary 2007). The body schema is as a suite of multimodal, action-oriented representations automatically subserving an agent's sensorimotor engagements with the world, e.g. by automatically changing the posture of our fingers when we grasp (Cardinali, *et al.* 2009b; de Vignemont *et al.* 2021). The neural basis of the body schema is highly plastic: even few minutes of repeated tool use are enough to modify it, allowing the tool to be included in the body schema, and to be used as if it were a part of our own body (Maravita and Iriki 2004; Martel *et al.* 2016; Romagno and Maravita 2021 for reviews). Since our bodily parts are typically phenomenally transparent while used (i.e. we do not need to monitor, control, reason about, or pay attention to the position of our fingers while grasping), then the tool will be phenomenally transparent too.[10]

It should now be obvious why (PTC) and Clark's proposal (sketched in §2) clash. Clark suggests that external resources are recruited into an agent's extended mind because they deliver high precision prediction

---

[10] The descriptions above hopefully clarify that (PT) has little to do with the claim that *experience* is transparent; i.e. that the only feature of experience we have introspective access to are features of the external mind-independent entities experience presents (cf. Tye 1995).

errors. So, according to PP, subjects *pay attention* to them. But we are intentionally directed to what we pay attention to, and indeed, what we pay attention to is typically "at the center" of our consciousness. So, if PP is correct, external, allegedly mind-extending resources, fail to satisfy (PTC) - thereby failing to qualify as mind-extending resources.

Yet, I will now argue, this should not lead us to abandon Clark's PP/extended mind hybrid. For, as I will show, (PTC) ought to be rejected: phenomenal transparency is neither a necessary nor a typical feature of mind-extending resources.

Notice: this is where the "switching of gears" happens. The two arguments I put forth to reject (PTC) are quite independent from PP. So the claim they support is not that *if* you like PP/extended mind hybrids *then* you should reject (PTC). The claim they support is that (PTC) *should be rejected full stop* - regardless of what one thinks about PP, extended cognition, and their relation.

Is the fact that my arguments are not based on PP a bad thing? Does it weaken my point? These are puzzling questions: since when giving *independent reasons* (in this case, PP-independent) for a claim makes it *worse*? Indeed, it seems to me that exactly the opposite is true: my claim would be weaker if the reasons provided to reject (PTC) were internal to PP. In fact, were my arguments against (PTC) dependent on PP, not only their truth would depend on the truth of PP, but their *relevance* would be

quite limited: they would be relevant exclusively to PP/extended mind hybrids. Thus, putting forth PP-independent reasons against (PTC) allows my claim to be of general interest (or at least, *more* general interest than otherwise).

But, then, what role does PP play in the overall economy of the paper? *Why bother with PP, when the "meat" of the paper is PP-independent?* There are, I think, three reasons to "bother" with PP here. First: having PP "at hand" will show that my solution to Clark's problem is superior to another, PP-specific, one (see §6.4). The other two are purely *practical* reasons. One is that my arguments against (PTC) have been designed to rescue Clark's PP/extended mind hybrid. The other is that the extended mind is now evolving in a way that is extremely dependent on PP (cf. Gallagher 2017; Kirchhoff and Kiverstein 2020; Constant *et al.* 2019), and sometimes deeply dependent on the role expected precision plays in PP (Kirchhoff and Kiverstein 2019). Thus keeping my rejection of (PTC) "in contact" with the PP framework allows me to reach my *target* audience easily.

Now, with this meta-commentary at hand, let me return to the main argument of the paper.

4 - A direct argument against (PTC)

Here's my direct argument against (PTC):[11]

---

[11] Whilst our analysis (and the morals we draw from them) diverge significantly, Andrada (2019) proposes a very similar

(P1) Internal resources need not satisfy (PTC) in order to be counted as constituents of a subject's cognitive machinery

(P2) Internal and external resources must be treated even-handedly

(C) External resources need not satisfy (PTC) in order to be counted as constituents of a subject's cognitive machinery

The argument is simple: two premises, and a conclusion. Let me examine them.

4.1 - *Premise 1*: inner cognitive resources can violate (PTC)

Consider the following thought experiment (Dennett 1978: Ch. 17). Dennett imagines the Pentagon tasks him with retrieving a nuclear warhead buried under the Earth surface. Since the warhead emits a radiation that damages *only* neural tissues, Dennett is forced to undergo a complex surgical procedure: his brain is thus placed in a life-supporting vat equipped with radio transmitters, allowing the normal brain-body message passing to occur. In this way, Dennett's brain can "pilot" Dennett's body and retrive the warhead, while staying safe from the radiation.

Now, here's how Dennett describes the moments just after the surgery:

> "Still feeling lightheaded, I was helped over to the life-support vat. I peered through the glass. There, floating in what looked like ginger-ale, was undeniably a human brain, though it was almost covered with printed circuit chips, plastic tubules, electrodes, and other paraphernalia. "Is that mine?" I asked. "Hit the output transmitter switch there on the side of

---

argument.

the vat and see for yourself," the project director replied. I moved the switch to OFF, and immediately slumped, *groggy* and nauseated, into the arms of the technicians, one of whom kindly restored the switch to its ON position." (Dennett 1978: 312)

Dennett's brain is not transparent: Dennett is clearly intentionally and attentively directed to it - he *peers* through the glass. Dennett's brain flouts (PTC) -but this *surely does not* expel Dennett's brain from his cognitive system. The idea that Dennett's brain (when properly functioning) is not part of Dennett's cognitive system is implausible. Moreover, were that failure of transparency to push Dennett's brain out of his cognitive machinery, *what would be left to constitute said machinery*? There seems to be no other constituent. Thus, if this failure to comply with (PTC) were to expel Dennett's brain out of his cognitive system, Dennett's exertion of his own cognitive abilities (e.g. the ability to pay attention to his brain) would become *unexplainable*.

Hence, we ought to concede that Dennett's brain is part of Dennett's cognitive system even when it fails to be transparent. Internal resources can thus fail to satisfy (PTC) just as (P1) requires.

This defense of (P1) might strike one as *too weak*. Surely, we can *conceive* sci-fi scenarios in which inner constituents flout (PTC). But conceivability is not a good guide to *possibility*. Borrowing an example from (Seth 2021): it is easy to imagine a boeing 747 flying backwards. But

it surely would be hasty to conclude that Boeing 747 *really can* fly backwards. In fact, they cannot.

This worry is easily allayed: there are *real-world* cases in which internal resources fail to be phenomenally transparent. Consider *neuro-feedback devices* (cf. Clark 2015a). These are neuroimaging devices whose *sole purpose* is to enable the user to *consciously and voluntarily monitor, control, and pay attention to* their inner neural goings-on, or a subset thereof (see Thibault *et al.* 2015). So, there are actual, real-world cases in which a subject's internal cognitive resources fail to be phenomenally transparent to the subject without thereby being pushed outside the subject's cognitive machinery. Thus inner resources seem able to flout (PTC) *as a matter of fact.* This is more than enough to vindicate the kind of possibility (P1) requires.

One might also try to resist my argument for (P1) asserting that "in some sense" Dennett's brain (and/or the inner goings-on of the neurofeedback device user) still are transparent, because they allow for an effortless, pre-reflective interaction with the world. For example, considering the case of neuro-feedback usage, Wheeler writes:

> "[...] there is a sense in which those states and processes are both transparent and visible. After all, one is still experiencing the world through those states and processes, so the transparency condition is satisfied. It is just that the world thereby revealed contains those very states and processes as objects."
> (Wheeler 2019a: 863).

This phenomenological characterization of "transparent *and* non-transparent" resources strikes me as incoherent - at least insofar transparency and non-transparency are mutually exclusive properties. On the assumption that what we are intentionally directed towards is not phenomenally inconspicuous, it fails to be phenomenally transparent in the relevant sense captured by (PT) and discussed here. Of course, the internal states and processes we are attentively and consciously directed to can still be *procedurally* transparent (and typically are), but *procedural* transparency is not transparency in the relevant sense discussed here - it's just *one component of* (PT). But what if (PT) was not the relevant way to characterize the relevant property of *transparency* discussed here? This is an important question I will come back to later (§6.1) to keep the structure of my parity argument clean.

4.2 - Premise 2: the parity principle

I ground (P2) in the *parity principle*, arguably the theoretical core of extended cognition:

> **Parity Principle**: If, as we confront some task, a part of the world functions as a process which, were it done in the head, we would have no hesitation in recognizing as part of the cognitive process, then that part of the world is (so we claim) part of the cognitive process (Clark and Chalmers 1998: 8).

The principle is a veil-of-ignorance-style test aimed at silencing our bio-chauvinistic prejudices (Clark 2013a: 195). It asks us to treat inner and outer resources *even-handedly*: if the φ-ing of inner resources is

sufficient for them to be counted as a constituent of an agent's mind or cognitive processing, then φ-ing of *external* resources is sufficient too. Otherwise put: external resources are not called to comply with any condition over and above the ones inner resources are called to comply with.

Why accept the parity principle? Because neuro-chaivinism is not very plausible, and to deny the parity principle is to accept a form of neuro-chauvinism (cf. Clark 2008: Ch. 5). Neuro chauvinists claim that *neural stuff* is, for some mysterious-and-yet-to-be-clearly-articulated reason, *more suited* to cognize than other stuff (cf. Searle 1992). Only if such a view is accepted the idea that, to be properly counted as vehicles, neural and non-neural resources must clear different requirements *makes sense*. But such a view is typically *not* accepted, and even adversaries of the extended mind agree that inner and outer resources must clear the *same* bar in order to be *really* counted as vehicles (cf. Adams and Aizawa 2001).

One could try to resist the parity principle on the ground that it *is a functionalist* principle, and functionalism ought to be rejected (for whatever reason). Some defenders of "radical"embodied theses seem to think that this is the case (e.g. Chemero 2009; Hutto and Myin 2013; Gallagher 2017).[12] They seem to think that since functionalism does not

---

[12] Shapiro (e.g. Shapiro 2019) provides very different reasons for the same conclusion.

recognize the importance of bodily features (at least, not to the right extent) they must reject functionalism, and thus that they are not entitled to use the parity principle.

Yet, they are entitled to use it, because the parity principle has nothing particularly *functionalist* to it: it is just an *even handedness* principle. The φ-ing of inner and outer resources relevant to the application of the parity principle may, but need not be, something pertaining their *functional* properties.

This is especially important to notice in this context, given that phenomenal transparency is, in a relevant sense, a phenomenal property and one might dispute that phenomenal and functional properties can be treated in the same way (cf. Chalmers 1996). Thus, to give an example of an application of the parity principle with a phenomenal property, suppose, as *integrated information theory* claims, that conscious properties are identical to specific amounts of informational integration (cf. Oizumi *et al.* 2014; see also Seth 2021: 59-74; 252-255 for a more accessible introduction). Suppose an amount of informational integration x is sufficient to instantiate the conscious property Q. And suppose we discover some external prop generating x when used by a subject. We can surely apply the parity principle and conclude that the prop is an external

vehicle of *Q*: after all, *were x generated "in the head"*, we would say that whatever is generating it is a vehicle of *Q*![13]

Note: the last statement is a *counterfactual* statement ("*were x generated in the head...*"). This is important to notice for it undermines an important line of attack against the parity principle. According to this line of attack (cf. Menary 2007; Sutton 2010; Gallagher 2017; Kirchhoff and Kiverstein 2019) the parity principle is unfit to substantiate claims of cognitive (or otherwise mental) extension. This is because they read the parity principle as requiring *a real and substantial similarity* between inner and (candidate) outer resources, which never or almost never obtains. According to this reading, the parity principle works as follows: here's a (candidate) external resource R ϕ-ing; is there - as a matter of fact - any inner resource R^* ϕ-ing? If yes, then R is actually an external vehicle of cognition, otherwise it isn't. Whilst *at times* the parity principle is employed in this way (cf Vold 2015), its original formulation requires nothing of that sort (see Wheeler 2011; 2019b; Clark 2011; 2013: 215; Smart 2022). The original formulation is *clearly* counterfactual ("were it done in the head", see above). So, *actually the parity principle works as follows*: here's a (candidate) external resource R ϕ-ing; if there were an internal resource R^* ϕ-ing would we say that R^* is a vehicle of cognition?

---

[13] Notice that I'm not cheating: Integrated information theory definitely is not a functionalist theory of consciousness (Doerig *et al*. 2019).

If yes, then *R* is actually an external vehicle of cognition, otherwise it isn't. So, no *real actual similarity* between the inner and the outer is required - and the problems connected with it are avoided.

Summing up: the parity principle is here pertinent, is well suited to support claims of cognitive extension, does not "smuggle in" troubling metaphysical commitments (e.g. to functionalism), it is not contested (adversaries of the extended mind accept it) and to deny it it so assert an implausible and unattractive form of neuro-chauvinism. It seems we have good grounds to accept (P2).

<div align="center">***</div>

I've thus far argued in favor of (P1) and (P2). If I'm on the right track, then (C) is true, and external objects *need not* satisfy (PTC) in order to be properly counted as cogs in the thinking machinery.

Notice that, if my arguments are correct, this conclusion is *sufficient* to rescue Clark's extended mind/PP hybrid: if phenomenal transparency is *not needed*, then its absence ceases to be a problem. Yet, whilst sufficient, the rejection of (PTC) I've just argued for still leaves Clark's hybrid vulnerable to a problem. One could reject (PTC) while still holding that phenomenal transparency is a *typical and important* feature of cognitive vehicles. Thus, phenomenal transparency may work as *reliable indicator* of cognitive extension (cf. Wheeler 2005; Kiverstein and Farina 2012; Clark 2015a). This role as a *reliable indicator* is sufficient to put

pressure on Clark's hybrid. After all, as Clark sets up things, transparency *never* obtains. Thus, were transparency a reliable indicator, we should conclude that, if Clark's proposal is correct, the mind rarely, if ever, extends.

I don't think such a pessimistic conclusion is warranted.

5 - A direct argument against transparency as an indicator of cognitive extension

Here's an argument against the idea that transparency reliably indicates cognitive extensions:

> (P1) If phenomenal transparency is a reliable indicator of cognitive extension, then:
>> (a) *at least* in paradigmatic instances of extended cognition, the external resource is phenomenally transparent; and
>> (b) *at least* the paradigmatic cases of phenomenally transparent resource usage are instances of extended cognition
>
> (P2) Both (a) and (b) are false
>
> (C) By *modus tollens*, phenomenal transparency is not a reliable indicator of cognitive extension.

(P1) is easily justified. "Indication" names a relation of *reliable and robust* correlation among property, holding in a range of circumstances (cf. Dretske 1988; Shea 2018). Thus, if *p* indicates *q*, we should expect them to be typically co-instantiated. (P1) expresses this is the least demanding form possible; i.e. by demanding the two properties to be co-instantiated

only in paradigmatic cases, rather than in the majority of cases ("Typically" is ambiguous in this respect).

What then, about (P2)? Here's the reasons to think it's true.

5.1 - (a) is false

Consider the following paradigmatic cases of cognitive extension:

Paradigmatic case #1: *Tetris* players must arrange falling geometric figures (*zoids*) in rows. To do so, they can press various buttons to translate, rotate or drop the *zoids*. Researchers (Kirsh and Maglio 1994) found that players' button presses *are not* finalized just to move (translate and rotate) *zoids* to target locations to then drop them. Participants also rotated not fully-on-screen *zoids* to speed up their recognition; they rotated zoids numerous times to *see* where they fitted best, and translated them horizontally numerous times to control the desired horizontal alignment. Thus, by manipulating zoids, participants simplified and sped up cognitive processing significantly, simplifying the cognitive demands imposed by the videogame (see Clark and Chalmers 1998; Clark 2008: 70-74; Kirsh 2019).

Paradigmatic case #2: Consider the use of pen and paper to do math. While confronted with even a modestly complex mathematical problem (e.g. multiply 398 by 72) many of us break it down into smaller steps (e.g. by using the method of partial products). Each step consists of an easily computable mathematical operation, the result of which is stored on paper for further use. The agent is thus using the numerals written on paper as a temporary memory buffer, which can be accessed through a quick glance (see Wilson 1994; Menary 2015).

Paradigmatic case #3: Experimental subjects are tasked with copying a pattern of colored blocks, dragging blocks (one at a time) from a virtual storage to a virtual workspace, which (together with the target pattern) are constantly displayed on screen. Eye-tracking technology revealed that participants repeatedly gazed at the target pattern, both to encode information about *color*, when choosing a block from the storage, and to encode information about *location*, when dropping the selected block in the workspace. The target pattern thus

> functioned as an *extended memory*: rather than encoding the entirety of the target pattern, participants consulted it through a perceptuomotor routine encoding small snippets of information as they were needed to guide their performance (see Clark 2008: 11-13; Krickel 2020).

Cases #1-#3 are all paradigmatic instances of cognitive extension. If (a) is true, the mind extending external resources should be phenomenally transparent in all three cases. But they are not. Subjects in #1-#3 surely *see* in the ordinary sense of the term, the *zoids*, the numerals on paper, and the colored blocks. And what we see, in the ordinary sense of the term, is not phenomenally transparent: we are aware of what we see, and what we see in the ordinary sense of the term is typically what our thematic consciousness is directed at. (cf. Farkas 2019: 46-48).

Further: the subjects of #1-#3 are plausibly *paying attention* to the props. They most likely want to succeed in these tasks, and it seems right to say that our chances of cognitive success in such tasks *would* be hindered, weren't we paying attention to the falling *zoids*, the numeral on paper or the colored blocks respectively. But the objects our attention focuses upon surely are not phenomenally transparent.[14]

---

[14]Clark (2019: 268) seems to concede the point, at least insofar he claims we *can* pay attention to our cognitive resources. But, the example Clark offers when seemingly conceding my claim does not seem to substantiate my conclusion. Clark invites us to consider the fact that agentive attention can safely intervene in purely internal processes without thereby pushing the attended information outside the bounds of cognition. His example is this: suppose a subject summons the mental image of a pair of mustaches and pays attention to their shape; it seems entirely correct to say that the relevant mustache-concerning information is still part of the subject's cognitive circuitry. Yet, in the case Clark invites us to imagine, the subject is paying attention to the *contents*, rather *the vehicles*, of her own cognition. But the extended mind thesis is a thesis concerning the *vehicles* of cognition, and the phenomenal transparency requirement hinges over the *vehicles* of cognition, rather than their contents. So, if Clark's example is intended to show that we can pay attention to our cognitive vehicles, the example fails. As a consequence, it is not entirely clear whether Clark *actually* rejects the phenomenal transparency requirement. And in fact, a few pages later, Clark claims (Clark 2019: 281) he fully endorses Kirsh's (2019) criteria for cognitive, which *do* include (PTC) (Kirsh 2019: 131-132).

Moreover: if (PT) is right, then phenomenally transparent resources shift the subject-world interface: we would encounter the world *through them*, just as we normally *see through* our eyes (and glasses/contacts), touch through our hands, and think through our neurons. Yet, it nothing of that sort seems to happen in #1 to #3. It seems entirely correct to say that, when playing Tetris, doing math or copying colored patterns, we continue to encounter the world through our biological transducers. Zoids and colored patterns are things encountered in the world, rather than *means through which* the world is encountered.

These, I submit, are all compelling reasons to think (a) is false - and thus to conclude that transparency is not the reliable indicator it is supposed to be.

5.2 – (b) is false

Consider the following paradigmatic cases of transparent tool use:

> *Paradigmatic case #I*: A large number of neuropsychological findings (e.g. Iriki *et al*. 1996; Berti and Frassinetti 2000; Maravita and Iriki 2004; Cardinali *et al*. 2009a; 2012; Canzoneri *et al* 2013; Bruno *et al*. 2019) suggests that, at least in primates, tools are incorporated in the body schema after a short period of usage.[15] When this happens, the tool is used, in the same automatic manner in which bodily parts are "used": just as when

---

[15] Yet, it is nor very clear what this incorporation amounts to. Researchers seems to oscillate between suggesting that *the tool is represented within the body schema* (e.g. Cardinali, *et al*. 2009a) and suggesting that *the prior representation of the body is modified* so as to "fagocitate" the tool (e.g. Bruno *et al*. 2019). Notice that the two claims are not equivalent at least when it comes to what the body schema represents: in the first case, it also represents *an external object*, whereas in the second case it doesn't. Moreover, in at least some cases (Berti and Frassinetti 2000; Constantini *et al*. 2014) researchers *do not* actually claim that the object is incorporated in the body schema: they claim only that tool use shifts the boundary between near and far space. But such a boundary shifts even in cases in which a subject's body schema is presumably *not incorporating* any external objects, such as when a physical barrier impedes a subject's grasping actions (e.g. Costantini *et al*. 2010). For the sake of argument, I'm going to assume these are just neuropsychological *minutiae* with no impact whatsoever on the philosophical picture emerging from this body of empirical work.

grasping one does not need to monitor her own hand, but the graping just happens, so too, when using an incorporated tool, one does not need to monitor, control, or pay attention to it, as its usage "just happens".

Paradigmatic case #II: A number of phenomenologists, most notably Martin Heidegger (1927/1961), argues that a correct description of our conscious experience when using tools will paradoxically make little to no reference to these tools. This is because, when a tool is used, the user's consciousness is not intentionally directed at it as an object in the world. Rather, the user's consciousness is directed at the task the tool allows one to perform, and it is that task, rather than the tool, that occupies the focus of the user's awareness.

Cases #I and II are paradigmatic cases of transparent tool use. Are they also cases of cognitive extension?

The parity principle suggests a negative answer. Recall: to apply it, we should imagine a *counterfactual* scenario where the φ-ing of the allegedly mind extending resource *R* is carried out by its inner counterpart by *R**. If the φ-ing of *R** makes it a vehicle, then the φ-ing of *R* makes *R* a vehicle too.

Yet, in the case at hand, the relevant external props are things like hammers, rakes and pliers. These items "take as inputs" bodily modevents and "yield as outputs" physical modifications of a subject's material surroundings. It's hard, to say the least, to imagine that kind of φ-ing going on "in the head" (cf. Rupert 2009: 31). And whilst maybe we *could* imagine an inner component able to modify the subject's surroundings in the relevant way (e.g. a sort of neural rake protruding from the subject's

head), what we are imagining seems to behave just like a regular *limb engaged in a pragmatic action* - indeed, it would be a neurally realized effector! And, typically, effectors are not counted as constituents of a subject's cognitive machinery.

So, the in-the-head case (insofar it can be imagined) suggests that our counterfactual piece of neural circuitry *would not* be counted as a constituent of the subject's cognitive machinery. But then, by the parity principle, the in-the-world case won't count as a constituent either. Again, phenomenal transparency is revealed *not* to be a reliable indicator of extension.

6 - Addressing some worries

Phenomenal transparency is neither a necessary, nor a typical, feature of cases of cognitive extension. Hence Clark's PP/extended mind hybrid is safe. Or is it? Maybe I've misunderstood what transparency amounts to. Maybe there are glaring flaws in my argument. Or maybe there's a way to fuse PP and the extended mind without having to reject (PTC). Let me address these worries.[16]

6.1 - Getting transparent on transparency

One could object that I got transparency wrong, and that (PT) should be rejected. Why?

For starters, (PT) is binary. If (PT) is accepted, something is either

---

transparent or not. But maybe transparency is not binary in that way. Maybe it is graded - certain things may be *more transparent* than others. The dichotomy transparent/non-transparent seems reminiscent of (and it may seem grounded in) Dreyfus's (2002b; Dreyfus and Dreyfus 1986) dichotomy between "mindless experts" and "attentive beginners". But this simple picture is *false*[17]: experts *attend to*, in a specific way, to what they are doing; they're in no way mindless (cf. Sutton *et al* 2011; Montero 2016, Cappuccio 2017 for discussion). And whilst *sometimes* the mind of experts "goes blank" and they *just act*, between this extreme and the extreme constituted by the beginner having to think twice about every movement there is a rich spectrum of forms of mentality, each exhibiting a specific degree of transparency.

Probably all of the above is correct, but it is natural to interpret the form of transparency there discussed just as *procedural* transparency. Indeed, whereas degrees of procedural transparency are *obviously* correlated to one's skills and expertise, degrees of phenomenal transparency aren't. Neither beginners nor expert chess players *literally* encounter the world through chess pieces. And, whilst most of us surely are not expert carpenters, hammers and similar tools are paradigmatic cases of phenomenally transparent tools. Phenomenal transparency really seems binary as (PT) suggests - either we are thematically intentionally

---

[17] Notice Dreyfus would agree: he clearly acknowledges there are many intermediate steps between these two extremes.

directed towards something or we aren't, there seems to be no "in-between". One could wonder how *phenomenal transparency* can be so dichotomous while *procedural* transparency isn't, given that the latter is a component of the former. But there are many, entirely non-mysterious, all-or-nothing phenomena that have a continuous component. A neuron either fires or stays silent - but while the *discharge* of a neuron is all-or-nothing, *charge isn't*. A person is either an adult or not, but age is clearly continuous. So, there's no mystery here. Something more worth investigation is the *exact* relationship between phenomenal and procedural transparency. My intuitive idea is that there's a *threshold* of procedural transparency that constitutes *at least* an enabling condition for phenomenal transparency. But I will not elaborate this idea further because it is not central to this paper.

One might further contend that (PT) mischaracterizes transparency by presenting it as a sort of *invisibility or non-given-ness* to consciousness. For example, in §5 I've claimed that since we often *see* (in the ordinary sense of the term) mind extending props, they're not transparent. Why? manifestly, because they are visible and given to consciousness - or so the objection suggests. But (the objection continues) that is the *wrong* way to think about transparency. Transparent objects *are* given to consciousness. They're just not given *as objects*. Rather, they're given to consciousness in a very specific way, as opportunities for action (cf. Andrada 2019; 2020).

We encounter transparent objects *not as objects*, but as affordances. Since (PT) does little to capture this, it mischaracterizes transparency, and should be rejected.

Whilst I agree with the objector that transparency is not non-givenness, I do not think that (PT) mischaracterizes transparency: (PT) allows us to capture transparency as a form of *non-thematic givenness*, which is what the objection aims at (at least, if I understand it correctly). Indeed, I think that equating transparency to affordances (as the objection does) *fails* to capture transparency. Surely, not everything we perceive as affording an action is transparent in the relevant sense: if Gibson (1979) is right, I see my door affording passing through, my chair affording sitting, and my mug of coffee affording drinking. But none of these things seems transparent in the relevant sense. Indeed, experimental data suggest we *(almost) always* perceive what objects afford (cf. Ellis and Tucker 2000; Tucker and Ellis 2001). But objects are *rarely* encountered as transparent.

To be clear: I'm not suggesting that, in a run of the mill sensory encounter with an object, we *explicitly infer*, from the look or aspect the object offers, what it allows. I'm not suggesting that, *normally*, our "inner monologue" goes like this:

> "This is a chair looking thus-and-so. Given the relevant features of that look (and some prior knowledge about chairs), it must have certain properties $p_1...p_n$, including p_{420} which - given the features of my body, allows me to sit. Let me now issue the

motor command C_{81} to my knees…"

Presumably, *nobody* - not even the "classiciest" of cognitive scientists - has ever believed that that were the case (cf. Schlicht and Starzak 2021 on direct perception). What I'm saying is that we normally encounter affordances *via* regular, unremarkable acts of perception, which present us objects *as such*. And objects *as such* are not phenomenally transparent. Thus, affordances do not provide us a good model of phenomenal transparency. They are unable to supplant (PT).

6.2 - A disparity in transparency?

One may fear that the parity argument I offered in §4 is hopelessly misguided. For, when it comes to transparency, there is a *fundamental disparity* between inner and outer vehicles. Which is the disparity? Different critics point in different directions. I will consider two options:

(a) Unlike outer vehicles, inner vehicles *need not be incorporated* - they're always already parts of our body. Hence the notion of phenomenal transparency cannot be rightfully applied to them.

(b) Unlike external vehicles, inner vehicles can be considered both objectively (from a 3rd person point of view) and subjectively (from a 1st person point of view). These two different points of view yield different results when it comes to phenomenal transparency, making the notion effectively useless.

I think none of these two options leads us to a *fundamental* disparity blocking the parity argument offered in §4.

To start, both (a) and (b) point to a *factual dissimilarity* between inner and outer vehicles. They leverage it to claim that the two kinds of

vehicles are not *similar enough* to be on a par. But, as noted in §4.2, Factual dissimilarities between candidate external vehicles and actual inner ones do not block parity arguments. Parity arguments do not ask us to adjudicate whether candidate external vehicles are sufficiently similar to inner ones to be *really* deemed vehicles of cognition. Parity arguments ask us to evaluate *counterfactual* scenarios: if the φ-ing of a candidate external vehicle R was carried out in the head by R*, would we consider R* a real and genuine vehicle of cognition? If so, then R is a vehicle of cognition too (see Wheeler 2011; 2019b; Clark 2011; 2013: 215).

But even if a "factual similarity" between inner and outer were needed, the differences (a) and (b) mentions would not block my argument.

Option (a) suffers from two distinct problems. First, if "being incorporated" entails "being transparent" (as commonly assumed, see for example Clark 2008:37-39), then it seems false that inner vehicles are "always already" incorporated. For, in some cases, inner vehicles fail to be transparent, as argued in §4.2. Secondly, it is commonly assumed that in order for a vehicle to be transparent it must be represented within a subject's body schema. But that does not seem to be the case. The brain does not seem to *represent itself*. There is no "brain strip" representing the brain in the cortical homunculus (cf Kandel *et al.* 2012: 364), and, as far as I can see, there is no reason to believe the brain "self-represents" as part of the agent's active body in any other way. Of course, none of

this is to deny that inner vehicles are nomological parts of our physical bodies "from the *get go*". But if this is what (a) is pointing out, it's either trivial (no one disputes this) or it equivocates on the relevant notion of incorporation ("being represented in the body schema" *vs* "being a piece of the physical body").

Option (b) points to a different fundamental disparity. Whereas we can consider inner vehicles both objectively (from a 3rd person point of view) and subjectively (from a 1st person point of view), we can consider external vehicles only objectively.

I find it hard to see the force of (b), for it is not entirely clear what the difference between 3rd and 1st person point of view amounts to *in this case*. It can't just be the fact that inner vehicles are both objects in the world *and* means to encounter the world. For the same is true of a variety of external objects, from the blind person's cane to glasses and high-tech prosthetics. Nor can it be the fact that "there is something that feels like" when a certain internal vehicle is present, but not when external (putative) vehicles are present. For sure such (putative) external vehicles do change our phenomenal experiences (e.g. taking one's glasses off).[18] And yet, I don't see *how else* (b) could be interpreted.

---

[18] Note: claiming this does not commit me to the "extended consciousness" thesis (i.e. the claim that the machinery of consciousness is partially constituted by external resources). I'm not claiming that external resources *constitute* consciousness, only that they *influence* it.

6.3 - A problem with trust and glue?

One might further contend that the use I've here made of the parity principle clashes with other "classic" criteria used to adjudicate cases of cognitive extension; namely the "trust and glue" criteria (Clark 2010). According to these criteria (among other things), in order to be properly counted as an external constituent of a subject's mind, an external prop must deliver information which is *automatically endorsed* by a subject. The information delivered by an external vehicle must *bypass by default* epistemic scrutiny, in a way that closely resembles phenomenal transparency (cf. Clark 2015). Since non-transparent resources do not bypass by default our epistemic scrutiny, then they don't satisfy the "truest and glue" criteria and are not real external constituents of our minds.

This worry is grounded on a close connection between epistemic scrutiny and phenomenal transparency. But there is no such connection. Indeed, epistemic scrutiny and phenomenal transparency are *doubly dissociable*. Non-transparent sources can deliver automatically endorsed information. Dennett's brain in §4.1 is an example. When Dennett sees his brain, his brain is not transparent to him. But Dennet still automatically trusts what his brain "tells" him. Cases of non-automatic endorsement of information delivered by phenomenally transparent vehicles are even more mundane. Sometimes, we doubt what we see, or whether we remember

something correctly - whilst the vehicles of these processes are surely as transparent as it gets, the information they deliver is not immediately trusted.

6.4 - Computational Shallowness

One could worry I ended up rejecting (PTC) only because I've been too quick on PP. A more careful consideration of the computational architecture of PP would solve the problem with Clark's extended mind/PP hybrid without forcing us to reject (PTC). It would kill two birds with one stone. Or so the objection goes.

The objection is misguided for two reasons. First, it gives us no reason to think we *shouldn't* reject (PTC). The objection assumes that retaining a commitment to (PTC) is a *desideratum*. But what justifies the assumption? I do not see any positive justification for it. Worse still, in §§4 and 5 I gave explicit and independent reasons *not to* retain a commitment to (PTC). So, unless some support favoring the status of (PTC) as a desideratum is provided, we have compelling reasons *no to* fiddle with PP's detail to retain our commitment to (PTC).

Further, there's a problem with the "pro (PTC)" proposal on the table. The proposal highlights something I only briefly hinted at in §2; namely the fact that generative models are *hierarchical*. Basically, this means that the computational mechanism instantiating the generative model depicts regularities in the world situated at different timescales, each

time-scale being a hierarchical level (Friston 2008). Lower level capture *fast* regularities with many fine-grained details (e.g. changes of sensory stimulations due to movement) whereas high level regularities capture *slow,* "abstract" regularities (e.g. the cycle of seasons). Now, the proposal (if I understand the objection correctly) is roughly this: Clark's PP/extended mind hybrid requires high precision only at *higher* levels. In order for us to cognitively engage with our surroundings, we must just have strong (highly precise) expectations about us engaging cognitively with our surroundings (cf Hohwy 2016).[19] We need not have highly precise expectations about the minute details about our sensory stimulation (low-level precision). And since *only* low-level precision disrupts skilled action impeding phenomenal transparency (cf Cappuccio *et al* 2020), PP and the extended mind can be happily married without having to sacrifice (PTC).

The problem with this proposal, I think, is the following: it *seems false* that, when we engage cognitively with the environment, only expectations at relatively *high* levels have *high* precision. If the precision at lower levels were low (or "average"), we *wouldn't* pay attention to the incoming sensory influx, and the information carried by the sensory influx would be basically ignored by internal processing. But this is manifestly not the case. Not only we often *do* pay attention to the props allegedly extending

---

[19] Using some PP-specific jargon: we must only have a high precision policy at a relatively high level.

our minds (e.g. when doing math with pen and paper, we *do* pay attention to the numerals written on paper), it seems simply false to say that the information coming from an external prop allegedly extending our mind is "ignored" by internal processing. Indeed, it is precisely *because* that information cooperates with our internal processing in a variety of ways that the external resource may be considered as an external extension of the mind in the first place! But note that often that information reaches internal cognitive processes *via* our sensory channels, and so the sensory flux is certainly not *ignored*.

So, to sum up: whilst one *might*, "tinker" with the neurocomputational apparatus of PP to allow PP and the extended mind thesis to hybridize in a way that does not force us to give up (PTC), we lack any positive reason to engage in such a project. Furthermore, the kind of "tinkering" thus far proposed seems to run counter the extended mind thesis. So, as things stand, my proposed solution *via* the rejection of (PTC) is superior to the one offered by a deeper scrutiny of PP neurocomputational apparatus.

7 - Concluding remarks

Here, I've diagnosed a problem plaguing Clark's PP-extended mind hybrid. According to Clark, external resources are recruited in an agent's extended mind in virtue of the *precise* prediction errors they are expected to deliver. Given that precise prediction errors tend to go hand in hand with attention, it follows that, according to Clark, subjects will *pay*

attention to the external, allegedly mind-extending, resources they are using. But this violates (PTC): the widespread idea that genuinely mind-extending resources *must* be phenomenally transparent. After my diagnosis, I've proposed a cure, claiming that (PTC) should be rejected, as transparency is neither a necessary, nor even a typical, feature of the vehicles constituting an agent's mind.

References

Adams, F., & Aizawa, K. (2001). The bounds of cognition. *Philosophical Psychology, 14*(1), 43-64.

Adams, F., & Aizawa, K. (2008). *The Bounds of Cognition*. Oxford: Blackwell.

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

Andrada, G. (2019). Mind the notebook. *Synthese, 198*, 4689-4708.

Andrada, G. (2020). transparency and the phenomenology of extended cognition. *Limite - Revista Interdisciplinaria de Filosofia Y psychologya*. Accessed from https://revistalimite.cl/index.php/limite/article/view/232, last accessed 11/01/20222

Auvray, M., & Myin, E. (2009). Perception with compensatory devices: from sensory substitution to sensorimotor extension. *Cognitive Sciences, 33*(6), 1036-1058.

Berti, A., Frassinetti, F. (2000), When far becomes near: remapping of spae by tool use. *Journal of Cognitive Neuroscience, 12*(3), 415-420.

Bruineberg, J. (2018). Active inference and the primacy of the "I can". In T. Metzinger, W. Wiese (Eds.), *Philosophy and Predictive Processing*: 5. Frankfurt am Main, The MIND Group. https://doi.org/10.15502/9783958573062.

Bruineberg, J. *et al.* (2018). The anticipating brain is not a scientist. *Synthese, 195*(6), 2417-2444.

Bruno, V., *et al.* (2019). How tool use shapes body metric representation: evidence for motor training with and without robotic assistance. *Frontiers in Human Neuroscience,* 13: 299.

Canzoneri, E., *et al.* (2013). Tool-use reshapes the boundaries of body and peripersonal spae representation. *Experimental Brain Research, 228*(1), 25-42.

Cappuccio, M. (2017). Flow, choke and skill: the role of the non-conscious in sport performance. In Z. Radmann (ed.) *Before Consciosuness* (pp.246-283). UK: Imprint Academic.

Cardinali, L., *et al.* (2009a). Tool-use induces morphological updating of the body schema. *Current Biology, 19*(2), R478-R479.

Cardinali, L. *et al.* (2009b). Peripersonal space and body schema: two labels for the same concept?. *Brain Topography, 29*(3), 252-260.

Cardinali, L. *et al.* (2012). Grab an object with a tool and change your body: tool-use-dependent changes of body representation for action. *Experimental Brain Research, 218*(2), 259-271.

Chalmers, D. J. (1996). *The Conscious Mind.* New York: Oxford University Press.

Chalmers, D. (2008). Foreword. In A. Clark (2008). *Supersizing the Mind.* New York: Oxford University Press.

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

Clark, A. (2003). *Natural-Born Cyborgs.* New York: Oxford University press.

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: the extended mind, extended. In R. Menary (Ed.), *The Extended Mind* (pp. 43-66). Cambridge, MA.: The MIT Press.

Clark, A. (2011). Finding the mind. *Philosophical Studies, 152*(3), 447-461.

Clark, A. (2013a). *Mindware* (2ⁿᵈ ed.), New York: Oxford University press.

Clark, A. (2015a). What "extended me" knows. *Synthese, 192* (11), 3757-3775.

Clark, A. (2015b). Radical predictive processing. *The Southern Journal of Philosophy, 53*, 3-27.

Clark, A. (2015c). Embodied prediction. In T. Metzinger, J. Windt (Eds.), *Open Mind*: 7(T), Frankfurt am Main: The MIND Group. https://doi.org/10.15502/9783958570115.

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

Clark, A. (2017a). Busting out: predictive brains, embodied minds, and the puzzle of the evidentiary veil. *Nous, 51*(4), 727-753.

Clark, A. (2017b). How to knit your own Markov Blanket: resisting the second law with metamorphic minds. 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. (2019). Reply to critics. In M. Colombo, E. Irvine, M. Stapleton (Eds.), *Andy Clark and His Critics* (pp. 265-302). New York: Oxford University Press.

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

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

Constant, A., *et al.* (2021). Representation wars: enacting armistice though active inference. *Frontiers in Psychology, 11*: 3798.

Costantini, M., *et al.* (2010). Where does an object trigger an action? An investigation about the affordances in space. *Experimental Brain Research, 207*(1), 95-103.

Costantini, M. (2014). When a laser pen becomes a stick: remapping of space by tool-use observation in hemispatial neglect. *Experimental Brain Research, 232*(10), 3233-3241.

Dennett, D. (1978). *Brainstorms*. Cambridge, MA.: The MIT Press.

Dennett, D. (1991). *Consciousness Explained*. Cambridge, MA.: The MIT Press.

De Vignemont, F., *et al.* (2021). What is the body schema?. In Y. Ataria, S. Tanaka, S. Gallagher (Eds.), *Body Schema and Body Image - New Directions* (pp.3-18). New York: Oxford University Press.

Di Paolo, *et la* (20222) Laying down a forking path: incompatibilities between enaction and the Free-energy Principle. *Philosophy and the Mind Sciences*, https://doi.org/10.33735/phimisci.2022.9187

Doerig, A., *et al.* (2019). The unfolding argument: why IIT and other causal structure theories cannot explain consciousness. *Consciousness and Cognition, 72*, 49-59.

Downey, A. (2018). Predictive processing and the representation wars. *Synthese*, 195(12), 5115-5139.

Dretske, F. (1988). *Explaining Behavior*. Cambridge, MA.: The MIT Press.

Dreyfus, H. (2002). Refocusing the question: can there be skillful copying without propositional representations or brain representations?. *Phenomenology and the Cognitive Sciences*, 1(4), 431-425.

Dreyfus, H. (2002b). Intelligence without representation. *Phenomenology and The Cognitive Sciences*, *1*(4), 367-383.

Dreyfus, H. & Dreyfus, S. (1980). A five-stage model of the mental activities involved in directed skill acquisition. Operations Research Centre. University of California, Berkeley

Ellis, R., & Tucker, M. (2000). Micro-affordance: the potentiation of components of action by seen objects. *British Journal of Psychology, 91*(4), 451-471.

Facchin, M. (2021a). Predictive processing and anti-representationalism. *Synthese*, 199(3), 11609-11642.

Facchin, M. (2021b). Extended predictive minds: do Markov blankets matter? *Review of Philosophy and Psychology*, https://doi.org/10.1007/s13164-021-00607-9

Farkas, K. (2019). Extended mental features. In M. Colombo, E. Irvine, M. Stapleton (Eds.), *Andy Clark and His Critics* (pp. 44-56). New York: Oxford University Press.

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

Friston, K. (1994). Functional and effective connectivity in neuroimaging: a synthesis. *Human Brain Mapping, 2*(1-2), 56-78.

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

Friston, K. (2008). Hierarchical models in the brain. *PLoS Computational Biology, 4*(11), e1000211.

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

Friston, K. (2010). The free-energy principle: a unified brain theory?. *Nature Reviews Neuroscience, 11*(2), 127-138.

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

Friston, K. *et al.* (2012a). Dopamine, affordance and active inference. *PLoS Comput Biol, 8*(1): e1002327.

Friston, K., *et al.* (2016a). Active inference and learning. *Neuroscience and Biobehavioral Reviews, 68*, 862-879.

Friston, K., *et al.* (2016b). Active inference and epistemic value. *Cognitive Neuroscience, 6*(4), 187-214.

Gallagher, S. (2008). Are minimal representations still representations? *International Journal of Philosophical Studies, 16*(3), 3521-3669

Gallagher, S. (2017). *Enactivist Interventions*. New York: OUP

Gallagher, S. (2018). The extended mind: state of the question. *The Southern Journal of Philosophy, 56*(4), 421-447.

Gibson, J. J. (1977). *The Ecological Approach to Visual Perception.* Boston: Hugton Muffin.

Grush, R., & Springle, A. (2019). Agency, perception, space and subjectivity. *Phenomenology and the Cognitive Sciences, 18*(5), 799-818.

Heersmink, R. (2015). Dimensions of integration in embedded and extended cognitive systems. *Phenomenology and the Cognitive Sciences, 14*(3), 577-598.

Heidegger, M. (1927/1961). *Being and Time*, New York: Harper and Row.

Hohwy, J. (2012). Attention and conscious perception in the hypothesis testing brain. *Frontiersin Psychology*, 3:96.

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

Hohwy, J. (2016). The self-evidencing brain. *Nous*, *50*(2), 259-285.

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

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

Hutto, D., & Myin, E. (2013). *Radicalizing Enactivism*. Cambridge, MA.: The MIT Press.

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

Iriki, A., *et al.* (1996). Coding of modified body schema during tool use by macaque postcentral neurons. *Neuroreport*, *7*(14), 2315-2330.

Kandel, E., *et al.* (2012). *Principles of Neural Science* (5th edition). New York: MacGraw Hill.

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

Kirchhoff, M. D., & Kiverstein, J. (2019). *Extended Consciousness and Predictive Processing: a Third-Wave View*. NewYork: Routledge.

Kirchhoff, M. D., & Kiverstein, J. (2020). How to determine the boundaries of the mind: a Markov blanket proposal. *Synthese*, *198*(5), 4791-4810.

Kirsh, D. (2019). When is a mind extended?. In M. Colombo, E. Irvine, M. Stapleton (Eds.), *Andy Clark and His Critics* (pp. 128-142). New York: Oxford University Press.

Kirsh, D., & Maglio, P. (1994). On distinguishing epistemic from pragmatic actions. *Cognitive Science*, *18*(4), 513-549.

Kiverstein, J. (2018). Extended cognition. In A. Newn, 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.

Krickel, B. (2020). Extended cognition, the new mechanists' mutual manipulability criterion, and the challenge of trivial extendedness. *Mind and Language, 35*(4), 539-561.

Lyon, P. (2006). The Biogenic approach to cognition. *Cognitive Processing, 7*(1), 11-29.

Maravita, A., Iriki, A. (2004). Tools for the body (schema), *Trends in Cognitive Sciences, 8*(2), 79-86.

Markman, A. B., and E. Dietrich (2000a). In defense of representation. Cognitive Psychology, 40, 138–171.

Markman, A. B., and E. Dietrich (2000b). Extending the classical view of representation. Trends in Cognitive Sciences, 4, 70–75.

Martel, M. *et al.* (2016). Tool-use: an open window into body representation and its plasticity. *Cognitive Neuropsychology, 33*(1-2) 82-101.

Menary, R. (2007). *Cognitive Integration: Mind and Cognition Unbounded*. Basingstoke: Palgrave Macmillan.

Menary, R. (2015). Mathematical cognition - a case of enculturation. In T. Metzinger, J. M. Windt (Eds.), *Open MIND*, 25(T). Frankfurt am Main, The MIND Group. https://doi.org/10.15502/9783958570818.

Menary, R., & Gillett, A. J. (2020). Are Markov blankets real and does it matter? In *The Philosophy and Science of Predictive Processing*, ed. D. Merdoça, S. Curado, and S. Gouveia, 39–58. London: Blomsbury Academic.

Merleau-Ponty, M. (1945). *Phenomenology of Perception*. New York: Routledge.

Montero, B. (2016). *Thought in Action*. New York: Oxford University Press.

Nave, K., *et al.* (2020). Wilding the predictive brain. *Wiley Interdisciplinary Reviews: Cognitive Science, 11*(6): e1542

Nolfi, S. (2002). The Power and limits of reactive agents. *Neurocomputing.* 42(1-2), 119-145.

Oizumi, M. *et al.* (2014). From the phenomenology to the mechanisms of cosnciousness: Integrated Information Theory 3.0. *PLoS Comput. Biol. 10*(5): 1003588.

Pfeifer, J., & Bongard, J. (2007). *How the Body Shapes the way we think.* Cambrdige, MA.: The MIT Press.

Piredda, G., & di Francesco, M. (2020). Overcoming the past-endorsement criterion: towards a transparency-based mark of the mental. *Frontiers in Psychology, 11*:1278.

Putnam, H. (1960). minds and machines. In S. Hook (Ed.), *Dimensions of Mind* (pp. 148-179). New York: New York University Press.

Ramstead, M. J., *et al.* (2019). Is the free-energy principle a formal theory of semantics? *Entropy, 22(8),* 889.

Roberston, I, & Kirchhoff MD. (2019). Anticipatory action: active inference in embodied cognitive activity. *Journal of Consciousness Studies. 27*(3-4), 38-68.

Romagno, D., & Maravita, A. (2021). Plasticity and tool use in the body schema. In Y. Ataria, S. Tanaka, S. Gallagher (Eds.), *Body Schema and Body Image - New Directions* (pp.117-133). New York: Oxford University Press.

Rupert, R. (2009). *Cognitive Systems and the Extended Mind.* New York: Oxford University Press.

Schlicht, T., & Starzak, T. (2021). Prospects of enactivists approaches to intentionality and cognition. *Synthese, 198*(1), 89-113.

Searle, J. (1992). *The Rediscovery of the Mind.* Cambridge, MA.: The MIT Press.

Seth, A. (2014). A predictive processing theory of sensorimotor contingencies: explaining the puzzle of perceptual presence and its absence in synesthesia. *Cognitive Neuroscience, 5*(2), 97-118.

Seath, A. (2021). *Being You.* UK: Faber & Faber.

Seth, A. K., & Friston, K. (2016). Active interoceptive inference and the emotional brain. *Philosophical Transactions of the Royal Society B: Biological Sciences, 371*(1708), 20160007.

Shapiro, L. (2019). Matters of the flesh. In M. Colombo, E. Irvine, M. Stapleton (Eds.), *Andy Clark and His Critics* (pp. 69-80). New York: Oxford University Press.

Shea, N. (2018). *Representation in Cognitive Science*. Caambridge, MA.: The MIT Press.

Silberstein, M., & Chemero, A. (2011). Complexity and extended phenomenological-cognitive systems. *Topics in Cognitive Science*, 4(1), 35-50.

Smart, P. (2022). Towards a mechanistic account of extended cognition. *Philosophical Psychology*, https://doi.org/10.1080/09515089.2021.2023123

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

Sutton, J. (2010). Exograms and interdisciplinarity: history, the extended mind, and the civilizing process. In R. Menary (Ed.), *The Extended Mind*, (pp.189-225). Cambridge, MA.: The MIT Press.

Sutton, J., *et al.* (2011). Applying intelligence to the reflexes: embodied skills and habits between Dreysuf and Descartes. *BJSP, 42*(1), 78-103.

Thibault, R. T., *et al.* (2015). Neurofeedback, self-regulation, and brain imaging: Clinical science and fad in the service of mental disorders. *Psychotherapy and Psychosomatics, 84*(4), 193-207.

Thompson, E., & Stapleton, M. (2009). Making sense of sense-making: reflections on enactive and extended-mind theories. *Topoi, 28*(1), 23-30.

Tschantz, *et al.* (2020). Learning action oriented models through active inference. *PLOS Computational biology*, 16(4): e1007805

Tucker, M, & Ellis, R. (2001). The potentiation of grasp types during visual object categorization. *Visual Cognition, 8*(6), 769-800.

Tye, M. (1995). *Ten Problems of Consciousness*. Cambridge, MA.: The MIT Press.

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

Wheeler, M. (2005). *Reconstructing the Cognitive World: the next step*. Cambridge, MA.: The MIT Press.

Wheeler, M. (2011). In search of parity about clarity. *Philosophical Studies, 152*(3), 417-425.

Wheeler, M. (2019a). The reappearing tool: transparency, smart technology, and the extended mind. *AI & Society, 34*(4), 857-866.

Wheeler, M. (2019b). Breaking the waves: beyond parity and complementarity in the arguments for extended cognition. In M. Colombo, E. Irvine, M. Stapleton (Eds.), *Andy Clark and His Critics* (pp. 81-98). New York: Oxford University Press.

Williams D. (2017). Predictive processing and the representation wars. *Minds & Machines*.

Williams, D., & Colling, L. (2018). From symbols to icons. *Synthese*.

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