Leaky Levels and the Case for Proper Embodiment

Mog Stapleton

Abstract: In this chapter I present the thesis of *Proper Embodiment*: the claim that (at least some of) the details of our physiology matter to cognition and consciousness in a fundamental way. This thesis is composed of two sub-claims: (1) if we are to design, build, or evolve artificial systems that are cognitive in the way that we are, these systems will have to be internally embodied, and (2) the exploitation of the particular internal embodiment that allows systems to evolve solutions with greatly decreased computational complexity – and thus to be flexible and adaptive in the ways that are characteristic of cognitive systems – means that the orthodox distinction between algorithm and implementation is undermined. Evolved cognitive systems are therefore likely not to instantiate the distinction between phenomenology, algorithm, and implementation. The leaky levels evident in evolved cognitive systems motivate an extension of orthodox embodied cognitive science to the internal, affective, "gooey" realm that has so far only been embraced by those in the enactive tradition. This suggests that if we are to build artificial systems that will be genuinely cognitive they will have to incorporate parts of the gooey realm currently considered to be "mere implementation."

1. Orthodox Embodied Cognitive Science

While there are many conceptions of embodiment that are relevant to philosophy of mind and cognitive science (see for example, Clark 1999; Wilson 2002; Anderson 2003; Ziemke 2003; Shapiro 2007) a broad overarching characterization used both by philosophers and those in the other cognitive sciences, including but not limited to robotics, is that our problem solving abilities are much less a matter of internal processing of information that comes into the system than orthodox cognitive science used to assume. Susan Hurley characterized the way of thinking about how minds work in the pre-embodiment revolution era as "the classical sandwich" (Hurley 1998): the thing that is doing the important work – the thinking or cognizing – is sandwiched between perception, bringing the information in, and acting according to the results of the information processing. On this model, the body is there to do the sensing and acting. Other than this the non-neural body is there just to keep the parts of the brain that these cognitive processes supervene on/are identical with (depending on your philosophical take) alive. Embodied cognitive science rejected this picture arguing that cognition cannot be cleanly separated from perception and action, and that many of our problem solving abilities are more a matter of adjusting the system itself, phylogenetically,

ontogenetically, or in occurrent action, such that the information processing that we would assume to be required if we were to set about programming these abilities into artificial systems, is actually "offloaded" onto the system's morphology.

While the term "offloading" in this respect is useful to see the difference from the pre-embodiment approach it is somewhat misleading because it implies that the standard case is that computations are done by the brain, but that they *can* be done by other structures instead. It rather seems to be the opposite. For many abilities the standard case is that we utilize the morphology of the body, the possibilities for action that it gives to us, and other structures in the environment so that we never have to do it all in our head (see Clark 1997, 2008b, [2001] 2013) – indeed most of us could not even do the calculating required for most of the things we do purely in our head. The result of taking embodiment seriously in robotics, is that when designing artificial cognitive systems, instead of deciding what function was to be implemented and then designing a program to implement it in a particular robot body (a top-down approach) one could rather consider what kinds of creatures achieve these tasks, consider how their bodies allow them to do it, emulate parts of that embodiment in the artificial system, and then program minimally to enable the system to engage in the bodily behavior (a bottom-up approach) (see Pfeifer and Bongard 2006, for a detailed examination of these principles at work in artificial cognitive systems and robotics).

The lesson for philosophy of mind should be clear from considering this case: the mind is both simpler and more complex than we previously imagined. On the one hand, we do not do as much of the information crunching that – on the orthodox approach – we assumed we must to support the ways that we perceive, think, and act. On the other hand, parts of what we had categorized as the mental when we assumed that minds were/supervened on neural information processing, seem now to be located in places that we are not intuitively comfortable in thinking of providing the supervenience/realizing base for mindedness (Clark 1997, 2008b; Clark and Chalmers 1998).

The choice is then to say that those are not parts of cognitive processes after all (the real cognition goes on in the head) in which case one needs to come up with a "mark of the mental" to distinguish real cognitive processes from processes that play a merely causal (rather than constitutive) role in cognitive processing (Adams and Aizawa 2008). Or, to bite the bullet and acknowledge that if we accept functionalism as the orthodox approach did, not only is mind not identical with the brain, but it is also perfectly consistent that mind is realized not merely by the brain. Andy Clark expresses this view in terms of what he calls the Larger Mechanism Story (LMS) as follows:

> Aspects of body and world can, at times, be proper parts of larger mechanisms whose states and overall operating profile determine (or minimally, help determine) our mental states and properties. (Clark 2008a, 39)

Functionalism about the mind of course implies multiple realizability, which entails that, provided that there are other materials which can implement the necessary processes, the functions that give rise to mind need not be limited to

our biological brains. This means that not only might artificial brains be created, but that there is no principled reason for confining minded processes to brain processes (biological or artificial). Thus, work on embodied (and extended) cognition that comes through this tradition is – in principle – entailed by the functionalism that orthodox cognitive science also assumed (see Wheeler 2010). I therefore refer to this approach to embodied cognition (of which Andy Clark [e. g., Clark 1997, 2008b] is a paradigm propagator) as "orthodox embodied cognitive science."[1]

Hence, while it might seem at first surprising that orthodox embodied cognitive science says very little about the role of the physiological – and homeostatic – body in cognition we can see that this is because (1) orthodox assumptions about cognition held that minded processes supervened in some way on brain processes, and (2) functionalism entailed that not only could the brain processes that it was assumed that minded processes supervene on be multiply realized in different implementations but could also be *partially realized* by other structures of the brain in the human case (hence "offloading"). The very term "extended" (in addition to "offloading") indicates that this kind of embodiment is still working within the orthodox "brainbound" tradition (see Clark 2008b) – but extending it outwards by following the implications of functionalism to their logical conclusion.

We can therefore see why even though fans of orthodox embodied cognition recognize that the body is important to cognition and mindedness, they are nevertheless loathe to think that physiological processes in the body proper make a contribution other than a causal or modulatory one to cognition. Any contribution that they do think is made e. g., by emotions, is made in virtue of those processes being represented in the brain. So while extended emotions fit in to this orthodox embodiment picture, this is not in virtue of the bodily contribution to cognition but rather once again following the implications of functionalism for the brainbound world view: i. e., taking emotions as represented in the brain as the standard case, and then showing that the processes we assumed were done in the brain are actually done (or can actually be partially done) through e. g., body posture, gesture, interpersonal engagement, or coupling with or structuring the environment in certain ways.

Let us then assume the basic tenet of orthodox embodiment: that cognition is not (at least not always) brainbound. Is there any more philosophical work to be done by asking the following question: if we are to build an artificial system that is genuinely cognitive, will implementing all of the processes that LMS throws light upon, be enough?

My hypothesis is that it will not be enough. This position, in and of itself, is not original; it is the position taken by many in the radical embodied[2] and

---

[1] I previously termed this "traditional embodied cognitive science" (Stapleton 2013).
[2] I use the term 'radical embodiment' here following Clark's (1999) distinction between simple and radical embodiment. Enactivism is therefore one (but not the only) version of radical embodiment.

enactive camps (see in particular Cosmelli and Thompson 2010; Thompson and Cosmelli 2011, which my project is heavily influenced by). However, these positions often put themselves, or are put, at odds with the orthodox embodiment approach because they build on assumptions which are not shared by all camps, e. g., the rejection of the representational/computational theory of mind, or the premise that a particular kind of self-organizing and self-creating organization underpins the development of cognition in key ways (see Thompson 2007; Di Paolo 2005, 2009; Di Paolo and Thompson 2014). These approaches may be thought of as biological rather than computational for a number of reasons but at the very least because they take as their standard case biologically cognitive creatures and seek to extrapolate from there to what is required for cognition, rather than taking the standard case to be the analogy of the mind to the computer.

I am sympathetic to both camps (though my publication history reveals that my intuitions mesh rather more with those of the biologically inspired approaches). Of course, as each side holds assumptions that contradict those of the other, one cannot be a full member of both camps at the same time (see Thompson and Stapleton 2009, for a discussion of why enactivism is not the same as externalism). What I then want to do is to take the spirit (rather than the details) of both camps and consider what insights it gives us. How can we do this, when orthodox embodied cognitive science is built upon functionalism, and enactive approaches reject functionalism about the mind?

I suggest that the spirit of orthodox embodiment is expressed by Clark (2008a) in LMS. Although this is a functionalist principle it is minimally functionalist: it does not entail a representational or computational view of the mind, it merely points to a mechanistic supervenience/realizing base for the mind. That this kind of mechanistic approach to the mind is still a kind of functionalism can be seen in Clark's work on "microfunctionalism" where he argues that functionalism does not need to be identified with high level formal descriptions such as beliefs and desires, rather what is essential to functionalism is that the "structure not the stuff counts" (Clark 1989, 31).

That cognitive creatures are mechanistic in this minimal sense is generally accepted in cognitive science – by both the orthodox and the radical. Where sides differ is in answering the question of what the minimal set of mechanisms is that enables/realizes cognition i. e., which are the ones we need to implement in order to build a cognitive system. The orthodox embodiment story clearly pushes the boundaries of the Marrian algorithmic level towards – and into – the implementational level for morphological features (Clark 2013). Yet, as explained above in virtue of its roots in the orthodox (brainbound) tradition, this minimal base does not include the internal goings-on in the physiological body. The intuition behind this is presumably that anything that is important that goes on in the physiological body is represented in the brain and so a functionalization of the relevant processes in the brain will include any relevant information from the body proper.

This is where I argue that the orthodox embodiment story errs. Let us talk in the mechanistic terms that are accepted by both them and the radicals, and

argue that the minimal realizing system is not quite big enough yet: That it must include at least some mechanisms that go on in the biological body (both the non-neural body and parts of the neural body that are typically functionalized out) as proposed by Cosmelli and Thompson (2010) and Thompson and Cosmelli (2011) with their thesis of "dynamic entanglement" (see also Clark 2013, for a discussion of dynamic entanglement from the orthodox embodied perspective).

Here I outline a story[3] which I propose should be accepted by both orthodox and radical embodimenters. While those in the enactive traditions will not think the story presented here complete as a minimal base for cognition, they should accept that it is at least part of what they consider the minimal mechanistic base and not reject it as externalist rather than embodied (Thompson and Stapleton 2009). And, because the story does not rest upon the assumptions of the radical approaches that orthodox embodiment rejects, and because it is presented as an extension of the mechanistic story and the fluidity of the algorithmic/implementational distinction that lies at the heart of the orthodox embodiment approach, without contradicting any of its own assumptions, orthodox embodimenters should also accept this story.

2. Introducing Proper Embodiment

The thesis of "Proper Embodiment" presented here is *that (at least some of) the details of our physiology matter to cognition and consciousness in a fundamental way* such that (at least some of) the mechanisms of cognition are so fine-grained that specifying the algorithm for cognition would entail specifying parts of the internal body normally considered to be background or enabling conditions for cognition.

I argue for this thesis through two independent theses: *internal embodiment* and *particular embodiment*. "Internal embodiment" is the thesis that the internal "gooey" body matters to cognition and consciousness in a fundamental way. "Particular embodiment" is the thesis that the particular details of our implementation matter to cognition. Taken together, these generate what I think is a compelling case that cognition is not merely embodied in the sense of orthodox embodied cognitive science, but Properly Embodied.

---

[3] The work presented here is a "big picture" view of the project developed in detail in my doctoral thesis (Stapleton 2012) situating this in respect to traditional and radical embodied cognitive science.

3. Internal Embodiment

Internal embodiment: the internal "gooey" body matters to cognition and consciousness in a fundamental way

In arguing for internal embodiment I focus on the role that interoception, the sense of the internal body, plays in cognition and consciousness. The term 'interoception' was originally used by Sherrington (1948) to refer to the sense of the visceral body (e. g., afferent information from smooth muscles and exocrine glands). A. D. Craig has since argued that due to sharing a common pathway through the spinal cord and processing areas in the brain, pain, temperature, and light touch should also come under the category of interoceptive senses and so "interoception should be redefined as the sense of the physiological condition of the entire body not just the viscera" (Craig 2002, 655). This sense of the physiological condition of the body gives a broad sense of how the body is faring. Although much of this information does not necessarily make it to conscious awareness, indeed Craig proposes that it is only in primates that this information is represented[4] in the right anterior insula, which is correlated with the sense of subjective feelings and emotions, it is nevertheless typically co-activated with the limbic motor cortex and so may underpin the motivational and valenced aspect of affective feelings as distinct from mere feelings of sensations. Interoception is therefore plausibly the basis for at least a minimal sense of value and thus intrinsic motivation, key parts of the cognitive apparatus that are underspecified by the orthodox embodiment paradigm but which a properly embodied story should give us an account of.

   Furthermore, recent work in affective neuroscience and predictive coding gives us reason to think that this interoceptive information may be involved in perceptual phenomenology. One such model, proposed by Barrett and Bar (2009) argues that when we perceive an object the brain makes a quick initial prediction about that object providing the gist of the situation but this does not yet correspond to our perception of the world. Rather, given this gist, the brain is left to predict the details of the situation based on previous knowledge, where "knowledge" is cashed out in terms of sensory-motor patterns that involve internal sensations including autonomic and endocrine information. On this model these predictions, and the filling out of the predictions, are recurrent and continue until the predictions at macro- and micro-levels no longer generate error signals when they are compared to incoming information. Information about internal bodily changes feeds in throughout this recurrency embedding affectivity into perception right from low level vision and including into the dorsal "where" visual stream.

   This model may initially seem unintuitive, influenced as we are by the Marrian framework of visual processing upon which, if affect plays any role it comes

---

   [4] I use this term in the (non-philosophically loaded) minimal sense understood in neuroscience.

in as an addition to fully formed perceptual contents. But consider an intuition pump from William James:

> Conceive yourself, if possible, suddenly stripped of all the emotion with which your world now inspires you, and try to imagine it as it exists, purely by itself, without your favourable or unfavourable, hopeful or apprehensive comment. It will be almost impossible for you to realize such a condition of negativity and deadness. No one portion of the universe would then have importance beyond another; and the whole collection of things and series of its events would be without significance, character, expression, or perspective. Whatever of value, interest, or meaning our respective worlds may appear embued with are thus pure gifts of the spectator's mind. (James 1902, 150)

While James appeals to emotions here, for him emotions are perceptions of bodily feelings, and so by definition available to conscious awareness, Barrett and Bar's model proposes that affect is playing an even more fundamental role in perception which they call "unconscious affect." They argue that:

> "Unconscious affect" (as it is called) is why a drink tastes delicious or is unappetizing … why we experience some people as nice and others as mean … and why some paintings are beautiful while others are ugly. (Barrett and Bar 2009, 1328)

This idea of "unconscious" contributions to experience that nevertheless shape the phenomenality of our experience is not unprecedented in philosophy. The phenomenological tradition has given us the concepts of pre-reflective and pre-intentional experience which gives all experience its characteristic 'colour' (see e. g., Ratcliffe 2010) but also contributes to the very structure of cognition. Ratcliffe (2005) for example draws on the phenomenological tradition to propose a reading of James' emotion theory that goes beyond emotions structuring our perceptual phenomenology to their being constituents of cognition. Understanding intentionality in the traditional phenomenological sense, as not merely the "aboutness" of a mental state but rather "conceptualized in practical terms, as an orientation that does not merely reveal but also differently configures the experienced world" (Ratcliffe 2005, 192) allows us to understand James as arguing that emotions/feelings are not only perceptions *of* bodily feelings but rather are constituted *by/through* both the perception of these bodily feelings and the feelings themselves.

While at first glance there might seem to be a tension here between on the one hand a part of an objective environment being revealed to one in virtue of one's senses and, on the other hand, one's world being a subjective construct, this tension is illusory. The claim is that there is an external world but we have access to only the parts of that world that are made available to us through our senses. What the phenomenological approach brings out – that the more biological approach may leave implicit – is that the senses do not make parts of that world available to us "as is" but rather the world is translated through our particular sensory mechanisms and possibilities for interaction such that experience is structured by these in a way we cannot eliminate. Thus, given that affect is intimately bound with our sensory capacities, it also shapes how we experience the world – "our world," and how we can act in that world. And, it is this claim that

if affect shapes how we experience the world then it also shapes how we can act in that world, that I take to be the heart of Ratcliffe's Jamesian/phenomenological claim that affect is constitutive of cognition. I have argued elsewhere in detail that drawing on the biological details of the interoceptive underpinnings of affectivity can give us good reason to think of affect as constitutive of cognition also in a non-phenomenological sense (Stapleton 2012). For the purposes of this chapter however the work outlined so far should be enough to motivate the plausibility of the weaker claim that in natural cognitive systems like ourselves, having an internal body shapes consciousness and cognition even when the interoceptive/affective information is unconscious/pre-reflective. And, because the information that feeds into cognition and consciousness is imbued with a natural value, in terms of value to the physiological system, to create an artificial system that is genuinely cognitive – and therefore has its own intrinsic values as a basis for motivation – we may need to implement some kind of functionally equivalent "internal body."

Internal embodiment – that *the internal "gooey" body matters to cognition and consciousness in a fundamental way* – on its own does not require a modification of Clark's LMS so much as it is an extension of it inwards. It contributes to the story something that was lacking in the standard functionalist framework: value and motivation, and begins to reintegrate the phenomenological with the functional to more properly address our actual explanandum in cognitive science: natural cognitive systems. Does Internal Embodiment on its own however actually require an internal body, even a functionalized version of one? It is not immediately obvious that the functions that the internal body plays in contributing to value and consciousness couldn't be implemented in the brain (or externally). After all, the orthodox embodimenter would argue, even in the biological case, the real contribution that they give to cognition and consciousness is in virtue of their representations in the brain. If this is the case, then it is not that the gooey body matters to cognition and consciousness in a fundamental way but rather that our gooey bodies implement functions that matter to cognition and consciousness in a fundamental way. While this may be an important addition to the orthodox embodiment story, it is nevertheless a trivial kind of *internal* embodiment, because the internality is not what is playing the key functional role.

Do we have reason for thinking that the internal, gooey body that has evolved as part of us, has a fundamentally more important role than a mere functional one at this Larger Mechanistic level? Or, to phrase this in different terms, are our physiological processes a mere happenstance of our evolution the essential functions of which can be happily implemented in a variety of materials and locations? I propose that this is not the case. Rather, natural cognitive systems are not only internally embodied but also embodied in a *particular* way that means large mechanistic functionalization of these processes just may not suffice[5]. This is the thesis of particular embodiment.

---

[5] For an extended meditation upon this theme in respect to creature consciousness that explicates the tight "entanglement" of the neurophysiological details, see Cosmelli and Thompson (2010); Thompson and Cosmelli (2011).

4. Particular Embodiment

Particular Embodiment: the particular details
of our implementation matter to cognition

Orthodox embodied cognitive science rests upon a version of functionalism that is expressed by Clark's LMS discussed above (see Wheeler 2010). Like most versions of functionalism this abstracts from the details of implementation because, as Clark puts it, the "structure not the stuff counts" (Clark 1989, 31). Any mechanistic view of the mind will of course endorse the principle that it is the structure and not the stuff that counts when it comes to bringing about cognitive processes. That there really is a distinction between structure and stuff, however, may not be as obvious as it first appears.[6]

   The thesis of particular embodiment is that the particular details of our implementation matter to cognition and hence any functionalization of the substructure of cognition would need to be at a fineness of grain that functionalizes these details. In order to motivate this thesis I will here put forward two "proofs of concept" drawn from evolutionary robotics: GasNets and evolved hardware.

5. GasNets

The principle behind evolutionary robotics is that, by emulating variation, heritability and natural selection one can artificially "evolve" robotic (or simulated) agents with complex behavior, gaining the standard advantages of neural nets, such as graceful degradation, as well as the targeted behavioral outcomes normally achieved through traditional programming. This is done by hooking up a group of neural networks to a task environment, or a simulation thereof, and selecting the most successful ones based on whatever fitness function you are using (i. e., those that are most successful – or least bad – at the task assigned). In order to increase variation of "genes" a few of those who were not most successful, but close by, are added and this group allowed to multiply while the rest are culled and recombinations of these "genotypes" and mutations are introduced. These steps are then repeated over and over, through many generations until networks evolve that can solve the task (the amount of generations needed to evolve a successful solution means that simulations are more practical than evolving networks using physical robotic agents at each stage).

   It has been known for some time that communication in the brain is not only mediated by electrical and chemical signaling but also gasotransmission, such

---

[6] The violation of this distinction is especially evident in autopoietic organization (Varela, Maturana, and Uribe 1974; Maturana and Varela 1980; Thompson 2007; for an accessible introduction, see Di Paolo and Thompson 2014). However, here I am concerned with putting forward a position that does not require a commitment to grounding cognition in autopoietic organization.

as through gases like nitric oxide (NO), carbon monoxide (CO), and hydrogen sulfide (H_2S). The assumption has always been, I take it, that in natural cognitive systems our implementation is gooey and complex as a result of our evolution but that this messy natural "design" could be abstracted away from and functionalized, and perhaps even improved upon. In short, there has reigned a culture of "electrical chauvinism" where it has been assumed that all of the important properties of cognition are represented at this level and molecular signaling and other gooey implementation can be factored out.

Smith and colleagues (2002) set up an experiment to compare the evolvability and adaptivity of solutions in standard artificial neural network models designed to model electrical transmission between nodes (NoGas) and an adaptation of the standard artificial neural network that was designed to also model gasotransmission (GasNet). The difference between the NoGas and the GasNet is that in the GasNet activations of nodes are not only a function of the inputs of the connected nodes (as with standard neural networks) but is also a function of the concentration of gas at that node.

The task that Smith and colleagues set was for robotic agents starting from an arbitrary position in a black-walled arena to find and navigate towards a white triangle while avoiding a white square. They show that basing their evolution of solutions on the GasNet class consistently produced successful solutions in fewer generations than evolution of solutions on the NoGas class. They argue that the GasNet solutions seem to be more evolvable because they are more amenable to being tuned to the particular characteristics of the environment, which is to say that the solutions are more flexibly adaptive. This adaptivity seems to arise from particular features of the gas diffusion mechanism which enable functions to be based on input patterns over time, which in turn allows noisy input to be filtered out.

This example from evolutionary robotics shows that a particular (gooey) feature of our embodiment plays a key role in evolvability leading to populations that can quickly adapt to a learning task and a particular environment. This is interesting in and of itself, but what is of particular relevance to us here is that both the GasNet and the NoGas controllers evolved functionally equivalent timing mechanisms. And yet, despite the functional equivalence in terms of the success criterion – the particular implementations resulted in a quite different amenability to being tuned to a particular environment. The moral for us to take from GasNets is that just because both GasNet and NoGas are successful solutions to the environment that they have evolved for (once they have evolved and reached 100 % fitness) and are therefore functionally equivalent in regard to the success criterion, this does not mean that the level of explanation at which we see the functional equivalence is the correct one to understand what is really key to the ability of each controller to succeed. That is to say what is key to a controller's being flexible and adaptive – the qualities that we are interested in if natural cognition is our explanandum – is not the same as what is key to the mere successful implementation of a function. By looking at the ease of evolvability and the mechanisms which underpin this amenability to being tuned to a

particular environment we can see that the relevant level of explanation for the adaptive behavior of the controllers is that which specifies the interaction of the gas and the nodes. The key point is this: In evolved systems this is not just the implementation, but rather it is the relevant level for the algorithm of an adaptive system. This is relevant not only over evolutionary timescales but, as Philippedes and colleagues (2005) note, also at the time scale of the (neurally plastic) changes themselves so that the biology of gas diffusion in real brains, and their subsequent modelling of GasNets parallels the embodied cognition approach to cognitive science, but internally. They state:

> In highlighting the functional importance of brain morphology, these phenomena take us increasingly further away from connectionist ideas and suggest that Pfeifer's notion of ecological balance, which requires a harmonious relationship between an agents' morphology, materials and control, can perhaps be taken inside the head. (Philippedes et al. 2005, 145)

This suggests that when it comes to cognition that functional equivalence may have to be at a much lower level than that specified by the LMS (and other functionalist approaches).[7]

6. Evolved Hardware

A second line of evidence for the thesis of particular embodiment also comes from one of the authors of the GasNet study: Adrian Thompson. While typically in evolutionary robotics algorithms are evolved in simulation and then transferred to hardware, this study used evolutionary algorithms to configure the switches on a Field-Programmable Gate Array evaluating the circuit based on its performance in the real-world (Thompson 1997). The aim of the experiment was to evolve a recurrent network of logic gates, and the Field-Programmable Gate Array is a digital chip which should therefore be ideally suited to this task. The surprising solution which evolved however was not based on logic gates. That is, the gates in the chip were not used to *do* logic. The solution that evolved exploited physical characteristics of the chip and behaviors that emerged. For example, a quarter of the cells in the array were clearly contributing to the target behavior as disabling them resulted in loss of the solution, but some of these cells were not even connected to the main part of the circuit. This defies the standard separation of algorithm and implementation. In this case the exploitation of physical characteristics of the chip enabled the system to evolve solutions which had greatly decreased computational complexity compared to traditionally designed algorithms. In this respect this example corroborates and strengthens the conclusions from the Gas-

---

[7] Note that this is not the same claim as is used against the parity principle in the extended mind debate. I am not here concerned with whether mental states such as beliefs or memories can be specified at a high or low functional level or whether implementational differences in these would violate their claim to instantiate these mental kinds. Rather, I am concerned with the substructure of the flexible adaptive behavior that enables/realizes cognition.

Net study. But this example from evolved hardware does even more than this: it gives us a real case where we can see that in evolved systems the line between algorithm and implementation is blurred so that it is no longer a trivial matter to implement an algorithm evolved on one particular piece of hardware on another piece of hardware in the way that functionalism assumes that one must.

7. The Case for Proper Embodiment

I have argued for two theses: internal embodiment and particular embodiment:
1. *Internal embodiment: the internal "gooey" body matters to cognition and consciousness in a fundamental way*
2. *Particular Embodiment: the particular details of our implementation matter to cognition*

The examples I have outlined in support of the thesis of particular embodiment give us good reason to think that the solutions that have evolved to make us the flexible, adaptive, neurally plastic cognitive systems that we are, are likely a result of the exploitation of our particular embodiment, both over evolutionary and developmental time, but very plausibly also over the time scale of the plastic changes that underpin new learning in hour-to-hour and day-to-day contexts. From this perspective the clean levels inherited from orthodox cognitive science and which remain implicit in orthodox embodied cognitive science – algorithmic and implementational – are revealed to be leaky in evolved systems.

While on its own the thesis of particular embodiment could be considered a mere extension to orthodox embodied cognition (see for example Clark's considerations of A. Thompson's work in Clark 2013), in combination with the thesis of internal embodiment it packs a much heftier punch: the internal physiological realm that interoception brings information from is a complex, dynamic system, and the lessons that we gain from the evolutionary robotics and hardware examples give us reason to think that it will not be easy to separate the algorithm of the relevant processes from their gooey implementation. The thesis of particular embodiment, while consistent with orthodox embodied cognitive science, is that much more radical in combination with the thesis of internal embodiment as together they not only suggest, as Philippedes and colleagues (2005) say, that the balance between morphology, materials and control can be taken inside the head, but that it can *also be taken into the body proper*.

The combination of internal embodiment and particular embodiment may seem to undermine orthodox embodied cognitive science because by pushing the leakiness of the algorithm/implementation boundary so far it is no longer clear whether there is a boundary at all.[8] However, while these theses may well

---

[8] Note that this position is not concerned with undermining extended functionalist positions of the kind that take as their explanandum mental states such as beliefs, desires, memories, etc. This is a thesis about the substructure of cognition conceived of as flexible adaptive behavior

contribute to undermining orthodox embodied cognitive science in combination with other assumptions or arguments (as for example is done in the enactive literature), as they are presented here, they do not need to be in conflict with the spirit of orthodox embodiment approaches. Taking this spirit to be accurately expressed by the LMS, the theses presented here motivate the modification of this to a "Smaller Mechanism Story" or a "nanofunctionalist" explanation (Stapleton 2012) where the processes that make up the substructure of cognition are much closer to the implementational details than traditionally envisioned. Being so close to the implementational details means that much of the body that was factored out on the orthodox approach is now going to play a role in the algorithmic substructure of cognitive processes. Nevertheless, a Properly Embodied cognitive science is not a biologically chauvinist position; a "smaller mechanism" or "nanofunctionalist" story implies that these functions could in principle be instantiated in different materials and could therefore in principle be extended – or rather external elements could in principle be "incorporated" (Clark 2008b; Thompson and Stapleton 2009) – but this instantiation is going to be at a much finer grain than traditionally assumed.[9] Proper Embodiment can thus be taken as extending orthodox embodied cognition inwardly, and thereby also extending the explanandum beyond abstract cognitive processes that are the target of much of cognitive science research back to the flexible, adaptive processes at work in evolved cognitive systems.

Bibliography

Adams, F. and K. Aizawa. 2008. *The Bounds of Cognition*. New York: John Wiley & Sons.

Anderson, M. L. 2003. Embodied Cognition: A Field Guide. *Artificial Intelligence* 149 (1): 91–130.

Barrett, L. F. and M. Bar. 2009. See It With Feeling: Affective Predictions During Object Perception. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* 364 (1521): 1325–1334.

Clark, A. 1989. Microfunctionalism: Connectionism and the Scientific Explanation of Mental States. Retrieved July 17, 2011, from http://www.era.lib.ed.ac.uk/handle/1842/1332.

–. 1997. *Being There: Putting Brain, Body and World Together Again*. Cambridge, MA: MIT Press.

–. 1999. An Embodied Cognitive Science? *Trends in Cognitive Sciences* 3 (9): 345–351.

–. 2008a. Pressing the Flesh: A Tension in the Study of the Embodied, Embedded Mind? *Philosophy and Phenomenological Research* 86 (1): 37–59.

–. 2008b. *Supersizing the Mind: Embodiment, Action, and Cognitive Extension*. Oxford: Oxford University Press.

–. (2001) 2013. *Mindware: An Introduction to the Philosophy of Cognitive Science*. Oxford: Oxford University Press.

Clark, A. and D. J. Chalmers. 1998. The Extended Mind. *Analysis* 58: 7–19.

---

that gives rise to the kinds of experiences that we then categorize according to these "cognitive" categories.

[9] Whether or not the possibility of being instantiated in different materials implies that nanofunctionalism entails multiple realizability is a question for another time.

Cosmelli, D. and E. Thompson. 2010. Embodiment or Envatment? Reflections on the Bodily Basis of Consciousness. In *Enaction: Towards a New Paradigm for Cognitive Science*, ed. J. Stewart, O. Gapenne, and E. A. Di Paolo, 361–385. Cambridge, MA: MIT Press.

Craig, A. D. 2002. How Do You Feel? Interoception: The Sense of the Physiological Condition of the Body. *Nature Reviews Neuroscience* 3 (8): 655–666.

Di Paolo, E. 2005. Autopoiesis, Adaptivity, Teleology, Agency. *Phenomenology and the Cognitive Sciences* 4: 429–452.

–. 2009. Extended Life. *Topoi* 28 (1): 9–21.

Di Paolo, E. and E. Thompson. 2014. The Enactive Approach. In *The Routledge Handbook of Embodied Cognition*, ed. L. Shapiro, 68–78. London: Routledge, Chapman & Hall.

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

James, W. 1902. *The Varieties of Religious Experience: A Study in Human Nature*. Bombay: Longmans, Green & co.

Maturana, H. R. and F. G. Varela. 1980. *Autopoiesis and Cognition: The Realization of the Living*. Dordrecht: Springer.

Pfeifer, R. and J. Bongard. 2006. *How the Body Shapes the Way We Think: A New View of Intelligence*. Cambridge, MA: MIT Press.

Philippides, A., P. Husband, T. Smith, and M. O'Shea. 2005. Flexible Couplings: Diffusing Neuromodulators and Adaptive Robotics. *Artificial Life* 11 (1–2): 139–160.

Ratcliffe, M. 2005. William James on Emotion and Intentionality. *International Journal of Philosophical Studies* 13 (2): 179–202.

–. 2010. The Phenomenology of Mood and the Meaning of Life. In *Handbook of Philosophy of Emotion*, ed. P. Goldie, 349–371. Oxford: Oxford University Press.

Shapiro, L. 2007. The Embodied Cognition Research Programme. *Philosophy Compass* 2 (2): 338–346.

Sherrington, C. S. 1948. *The Integrative Action of the Nervous System*. Cambridge, UK: Cambridge University Press.

Smith, T., P. Husbands, A. Philippides, and M. O'Shea. 2002. Neuronal Plasticity and Temporal Adaptivity: GasNet Robot Control Networks. *Adaptive Behavior* 10 (3–4): 161–183.

Stapleton, M. 2013. Steps to a "Properly Embodied" Cognitive Science. *Cognitive Systems Research* 22–23: 1–11.

Stapleton, M. 2012. "Proper Embodiment: the Role of the Body in Affect and Cognition." Phd Thesis, University of Edinburgh.

Thompson, A. 1997. Artificial Evolution in the Physical World. In *Evolutionary Robotics: From Intelligent Robots to Artificial Life (ER'97)*, ed. T. Gomi, 101–125. Ontario: AAI Books.

Thompson, E. 2007. *Mind in Life: Biology, Phenomenology, and the Sciences of Mind*. Cambridge, MA: Harvard University Press.

Thompson, E. and D. Cosmelli. 2011. Brain in a Vat or Body in a World?: Brainbound Versus Enactive Views of Experience. *Philosophical Topics* 39 (1): 163–180.

Thompson, E. and M. Stapleton. 2009. Making Sense of Sense-Making: Reflections on Enactive and Extended Mind Theories. *Topoi* 28 (1): 23–30.

Varela, F. G., H. R. Maturana, and R. Uribe. 1974. Autopoiesis: The Organization of Living Systems, its Characterization and a Model. *Biosystems* 5 (4): 187–196.

Wheeler, M. 2010. In Defense of Extended Functionalism. In *The Extended Mind*, ed. R. Menary, 245–270. Cambridge, MA: MIT Press.

Wilson, M. 2002. Six Views of Embodied Cognition. *Psychonomic Bulletin & Review* 9 (4): 625–636.

Ziemke, T. 2003. What's That Thing Called Embodiment? In *Proceedings of the 25th Annual Meeting of the Cognitive Science Society*, ed. R. Alterman and D. Kirsh, 1305–1310. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.