Cognitive ontology in flux: the possibility of protean brains

Daniel D. Hutto [ID]*, Anco Peeters [ID] and Miguel Segundo-Ortin [ID]

School of Humanities and Social Inquiry, Faculty of Law, Humanities and the Arts, University of Wollongong, Wollongong, NSW, Australia

This paper motivates taking seriously the possibility that brains are basically protean: that they make use of neural structures in inventive, on-the-fly improvisations to suit circumstance and context. Accordingly, we should not always expect cognition to divide into functionally stable neural parts and pieces. We begin by reviewing recent work in cognitive ontology that highlights the inadequacy of traditional neuroscientific approaches when it comes to divining the function and structure of cognition. Cathy J. Price and Karl J. Friston, and Colin Klein identify the limitations of relying on forward and reverse inferences to cast light on the relation between cognitive functions and neural structures. There is reason to prefer Klein's approach to that of Price and Friston's. But Klein's approach is neurocentric - it assumes that we ought to look solely at neural contexts to fix cognitive ontology. Using recent work on mindreading as a case study, we motivate adopting a radically different approach to cognitive ontology. Promoting the Protean Brain Hypothesis, we posit the possibility that we may need to look beyond the brain when deciding which functions are being performed in acts of cognition and in understanding how the brain contributes to such acts by adapting to circumstance.

Keywords: enactivism; cognitive ontology; functionalism; neurodynamics; pluripotency

Introduction

Deciding on the right cognitive ontology is a matter of deciding what ultimately belongs in our theoretical toolbox when it comes to understanding the function and structure of cognition. Very broadly construed, it is a matter of deciding which taxonomic concepts or constructs we ought to employ "in the study of human thinking and acting" (Anderson 2014, 311; see also Klein 2012, 959).

The enterprise of articulating the right cognitive ontology, when taken up in a realistic spirit, is standardly conceived as one of carving cognition at its supposed natural, neural joints.[1] In line with classical cognitivist assumptions, there is a default tendency to understand the project of identifying the best cognitive ontology in an internalist, neurocentric frame. For our purposes, an approach to cognitive ontology counts as neurocentric if it only looks to assign cognitive functions to neural structures on the assumption that cognition is primarily or always restricted to neural structures.

Neurocentrism assumes that the whereabouts of the mechanisms responsible for completing cognitive tasks are always to be found solely and wholly within the brain. It is an unquestioned assumption, for many in cognitive science, that "cognition is ... realized in

---

*Corresponding author. Email: ddhutto@uow.edu.au

the brain alone" (Aizawa 2015, 756). This view is disputed, for example, by enactive views of cognition (see, e.g. Thompson 2007). Still, it is only with that neurocentric assumption in place, that one might consider it to be an "unremarkable truism", as Klein (2012) does, that, instead of the whole brain-body-environment system, "a cognitive task is always performed by a network of brain regions working in concert" (952).

Through a neurocentric lens, the aim of cognitive ontology is thus to map "mental functions onto their neural substrates" (Lenartowicz et al. 2010, 678). And, if such mappings are successful, the expectation is that the articulation of a correct cognitive ontology would yield an understanding of "how brain systems give rise to mental function" (Poldrack and Yarkoni 2016, 588).[2]

This paper motivates moving beyond neurocentrism by taking the Protean Brain Hypothesis, or PBH, seriously. The PBH conjectures that brain structures are functionally malleable: that they make use of neural structures in inventive, on-the-fly improvisations to suit circumstance and context. Hence, we should not always and everywhere expect brains to divide into natural, functionally stable parts and pieces. We argue that to deny this possibility at the outset is to foreclose on live possibilities and to invite the real danger of imposing an assumed cognitive ontology onto the brain rather than discovering it within a wider brain-body-environment system.

The paper divides into four main parts and develops as follows. Section one reviews reasons for preferring a Klein-style context-sensitive, bottom-up approach to cognitive ontology that makes use of data-driven meta-analyses of neuroimaging data to carve cognition at its supposed natural joints. Opting for a different kind of context-sensitive approach – one that goes further by looking to non-neural contexts to disambiguate cognitive functions – we introduce the possibility that brains may be protean. Section two seeks to motivate taking the PBH, seriously as a radically different approach to cognitive ontology. Reviewing recent work on mindreading as a case study, reasons are given for being open to the possibility that we may need to look beyond the brain when deciding which functions are being performed in acts of cognition and in understanding how the brain contributes to such acts. Section three presents Anderson's alternative approach to cognitive ontology as a way of taking the PBH seriously. Anderson's alternative considers cognitive ontology to be distributed among brain, body, and environment, at least *ab initio* and is thus situated within the larger landscape of the embodied, enactive and ecological movement in cognitive science.[3] Section four analyses a perceived tension in Anderson's approach: that of embracing enactivism while holding on to a functionalist understanding of cognition. It is argued that this tension disappears if functionalism is taken in its broadest, most minimal sense. It is concluded that, although a neurocentric cognitive ontology is not logically impossible, there are no strong reasons to expect it is in the cards. Moreover, there are sufficient reasons to consider opting for a distributed, enactive, dynamic vision of the brain's basic and more protean cognitive ontology.

1.   Carving up our cognitive ontology

The task of articulating a cognitive ontology has been characterized as that of identifying "the stock of basic functions and relations that we use to describe cognitive tasks" (Klein 2012, 954). Crucially, though not universally accepted, this way of rendering the cognitive ontology project lives by the presupposition that "function should predict the structure and conversely structure should predict function" (Price and Friston 2005, 262).

Divining the appropriate links between function and structure has not proved easy. As Klein (2012) explains, traditional cognitive ontologies – in particular those that depend on

naïve regional function attributions – are fatally flawed. Such approaches make use of complementary forward and reverse inferences.

The recipe for forward inference is as follows. First, set up an experiment where participants are presented with two tasks. Second, take increased activity in brain region X during task Y1 when compared to control baseline task Y2 as evidence that region X is involved in task Y1. Third, infer that X is performing a function Z to complete task Y1. Reverse inference, as the name indicates, goes the opposite way. First, assume – based on the results of forward inference – that function Z is performed by brain region X. Second, present participants with tasks Y3 . . . Yn. Third, conclude that during tasks with increased activity in region X, function Z is performed. In short, forward inference is used to attribute functions to brain regions, and reverse inference is used to discern whether certain functions are involved when specific tasks are performed (see also Anderson 2015, 71).

Forward and reverse inference aim to get at a cognitive ontology by inferring which brain regions perform a specific function and which tasks involve those functions. The problem with this simplistic approach, as Klein (2012) identifies, is that "brain regions appear to be pluripotent: that is, there is [a] many-to-one mapping between functions and brain regions" (953). For example, brain region X might be performing function Z, but there is no way to determine, using these methods alone, that it is not also performing different functions in different tasks. Though this problem is dubbed the "problem of reverse inference" (Price and Friston 2005), inferences in both directions are unsafe: as Klein observes, "arguably, pluripotency is equally troubling for the forward inference step" (Klein 2012, 953).

In sum, the moral is that though the inferred functional attributions are meant to illuminate brain structures and vice versa, it is hard to see how, using only these methods, neuroimaging pinpoints a function to a brain region, if it is unclear what that brain region does. And it is unclear what a region does, since the brain regions are put to different sorts of work in different contexts.

One way to attribute functions to brain regions is by "going general" and abstract from the type of function that is attributed. Consider, for example, the function of a hand. In enabling the drinking of morning coffee, it performs the function of grasping a cup. When judging the attractiveness of a fabric, it performs the function of testing its smoothness or texture. On seeing a friend, it performs the function of waving. A hand performs all of those functions and many more, but we ascribe it a quite general function – that of taking different configurations (cf. Klein 2012, 953–954). A similar move can be made with regard to brain regions, Price and Friston (2005) suggest.

The trouble is that the "going general" strategy comes at the cost of sacrificing explanatory power. Consequently, by going general we must give up "on answering a lot of questions that depend on the details" (Klein 2012, 955).

To avoid loss of explanatory power, Klein instead suggests a context-sensitive way of attributing functions. According to Klein's diagnosis, the ultimate source of the problem of reverse inference is that the function of brain regions is being "considered in isolation from the broader neural context" (Klein 2012, 952). Consider the hand again – its function differs: in the consuming coffee context; in the judging the qualities of a fabric context; in the social, greeting context. We can disambiguate its function in a context-sensitive manner. Klein argues we should make basically the same move when deciding on a brain region's function – namely, by looking at the larger context in which it operates. Importantly, however, Klein (2012) restricts his interest to only looking at a wider, *neural* context, understood as, "the overall network in which a region is participating" (Klein 2012, 957).

According to Klein, we can discover the functions of neural structures by looking to wider neural contexts and thus disambiguating the precise nature of cognitive tasks. Consider Klein's (2012) example of how to decide, in a context-sensitive way, which of two possible functions the pistons on diesel trucks are performing:

> *Which function the piston performs depends on things external to it:* whether it is powering or slowing the truck depends on the ignition system and the valve timing. Both of the context-specific functions of the piston are straightforward, easy to understand, and useful to cite. (Klein 2012, 955, emphasis added)

In the end, Klein's preferred approach is not to impose pre-conceived task ontologies on differential activity in brain networks, but – in the vein of Poldrack, Halchenko, and Hanson (2009) – to use data-driven meta-analyses of neuroimaging data to generate task ontologies that can differentiate activity in brain networks. In sum, to let the brain data generate its own discriminatory dimensions. Such methods can be used "to see whether tasks we think are similar (or distinct) activate similar (or distinct) networks. Evidence for similar activation is evidence that we have divided up tasks correctly" (Klein 2012, 957–958).

The end result of this bottom-up filtering process, if all goes well, is the discovery of "relatively stable" and "commonly activated" sets of regions. And, if we can get this sort of view of a whole network then, Klein tells us, "we can then drill back down into the function of specific areas to see how they look" (958). In other words: we have to zoom out of the functional structure before we can zoom back in.

An alternative approach – the PBH – assumes that neural structures are essentially mutable and can be enlisted for many kinds of cognitive tasks. Accordingly, PBH embraces the idea that neural pluripotency is not a bug but a feature. Thus, like Klein, the PBH recommends looking to context to disambiguate the cognitive contributions of neural structures in the performance of any given cognitive task. Yet it does so by going beyond the neural – namely, by using the system's environmentally situated activity to understand the contributions of neural activity. Concomitantly, discovering what some bout of neural activity within a given network or region is doing is to get clear about the nature of the world-involving cognitive task in which it is involved. On such a fluid vision of cognitive ontology, even if neuroimaging manages to pinpoint the similarities and differences in structural features of brain networks and regions across tasks, such discoveries will not, a fortiori, shed light on stable functions of neural structures.

2.   Protean brains and mindreading: a case study

Why take PBH seriously? We should take the hypothesis seriously to the extent that we do not want to foreclose on the possibility that the performance of cognitive tasks will not always neatly carve up into natural, neural joints – and that cognition may not always take place in a wholly neural context. Consider the experimental work using correlative neuroimaging studies devoted to identifying the existence of a dedicated network of brain regions that are reliably involved in so-called mindreading tasks (see Koster-Hale and Saxe 2013 for an overview). It has put forward two main candidates for mindreading networks: (1) cortical midline structures (CMS) – including medial pre-frontal cortex (MPFC), adjacent rostral anterior cingulate cortex (rACC) and medial posterior parietal cortex (MPPC) – and (2) the bilateral temporal parietal junction (TPJ), both left (L-TPJ) and right (R-TPJ).

Mahy, Moses, and Pfeifer (2014) supply reasons for thinking that the TPJ network is the stronger contender. Let us assume they are right. Can we thereby infer that this set of

regions is dedicated to mindreading in the brain? Immediately, we apparently face the problem of pluripotency. The TPJ network is not selectively deployed in mindreading, *and only* mindreading tasks. Commenting on this lack-of-specificity problem, Apperly (2011) reports the conclusions reached by Legrand and Ruby (2009) in their meta-review of neuroimaging studies: the so-called "mindreading network" was implicated not only in mindreading tasks but also in tasks involving "introspection, recalling information from memory, reasoning in general, and particularly inductive reasoning under conditions of vagueness and uncertainty" (62). Thus many tasks that do not appear to inherently require mindreading nevertheless evoke activity in the TPJ, thus violating the specificity criterion (see Mahy, Moses, and Pfeifer 2014, 71).

The above example shows that "the challenges posed by the many-to-many mapping between regions and functions is not dissolved by the network perspective" (Pessoa 2014, 400). An option at this juncture, focusing again on the style of solution offered by Price and Friston (2005), is to go abstract and to assign a more general function to the entire network – namely, to assume that its function is that of, say, integrative, abstract reasoning. Here Klein's (2012) reasons for avoiding such a move, based on loss of explanatory power, when it comes to thinking about the functions of regions, seems equally apt when it comes to thinking about the function of networks: looking for some overarching domain-general function of the whole network would likewise diminish explanatory power and risk vacuity (955).

In line with Klein's context-sensitive answer to the pluripotency problem about networks, Apperly (2011) sums up the current state of the art in social neuroscientific proposals about where mindreading putatively happens in the brain as follows:

> The overall picture that emerges is surprisingly consistent. Mindreading recruits a complex network of functional and neural processes. Included among these are neural regions that appear highly selective for mindreading, but it seems clear that mindreading *per se* is a function of the network rather than of specific, specialized brain regions. (59)

Apperly is in favour of assigning a dedicated mindreading function to the entire brain network, not to a single region. He is alive to the danger of over-emphasizing the importance of single regions – such as the right temporal–parietal junction (R-TPJ) – and assigning it the function of "mindreading". The problem is that doing so would likely result in excluding a broader set of neural regions that also serve important or even indispensable mindreading functions (Apperly 2011, 70). Apperly's assessment is clearly compatible with Klein's (2012) proposal of using data-driven meta-analyses of networks in order to try to determine the yet more specialized functional contributions of specific brain regions within neural networks (cf. Klein 2012, 959).

For example, it might be – as some have claimed – that the R-TPJ is a brain region that is selectively enlisted for tasks requiring the interpretatively complex attribution of mental states. Saxe and Wexler (2005) found that R-TPJ activity is increased compared to other regions when the professed beliefs or desires of story protagonists conflicted with subjects' background expectations about what these characters ought to believe or desire. Moreover, this region is not similarly recruited for other tasks that involve assessing other, more general, socially relevant facts about persons.

These proposals all proceed from the assumption that we ought to be trying to assign relatively stable functions to networks and to sub-regions in those networks – say, "mindreading" and "mental state attribution" respectively. Yet the need to and value of making such an assumption is challenged by certain ways of answering the hotly debated question

of what is actually going on when we make sense of others during the relevant acts of social cognition. The issue at stake in such debates is what form understanding another's action actually takes and what cognitive tasks it involves. Currently, it is not only far from clear which tasks are involved in mindreading, or indeed, whether we mindread at all but, more importantly, whether the performance of such tasks neatly carves up so that the underlying cognition takes place within a wholly neural context.

For example, it has been argued that in second-person contexts, we get by in understanding another's actions without making any mental state attributions at all (Gallagher 2001, 2008; Hutto 2004, 2008; Zahavi 2011). This is because we either manage to respond directly to the expressed attitude of another or because, as often happens, the other supplies an account of why they did what they did for our consumption, obviating the need to infer and attribute mental states. Indeed, it has been argued that

> in many everyday cases of sense-making, it suffices to attribute "factive" reasons—reasons in terms of normatively salient aspects of the (social) environment. So *what* interpreters are doing in these default cases should *not* be coined in terms of mindreading or mentalizing. (Strijbos and de Bruin 2012, 144)

Even more radically, it has been argued that, quite generally, we do not make mental state attributions when understanding people's actions; and, indeed, that "people do not employ a belief-desire psychology at all" (Ratcliffe 2009, 386).

Consequently, we may be fundamentally wrong in assuming that any kind of mindreading lies at the heart of our everyday way of understanding reasons for action.

> If we start out with an interest in mindreading there is a tendency to see a need for it in almost any social activity. How else could we explain an infant's ability to engage in a teasing interaction with its carer, a child's ability to understand everyday social interactions, or adults' remarkable ability to work out what one another are talking about? Surely in all these cases it is necessary to think about what other people know, think, want or intend? Actually, there are many reasons for thinking that this is often unnecessary. (Apperly 2011, 114)

Some maintain of these so-called phenomenological critiques – all of which question the centrality of standard mindreading accounts of how we make sense of everyday actions – that, even if they hold good, they only apply to our personal level understanding of the nature of the relevant social cognitive tasks (Herschbach 2008; Spaulding 2010). As such, this allegedly frees us up to go sub-personal where the interest is not in

> *what* it is that we do when we interpret each other, but *how* we do it (emphasis in the original). Answering this how-question involves uncovering the cognitive mechanisms and processes that underlie our sociocognitive capacities and cause our judgements about other people's reasons for action. (Strijbos and de Bruin 2012, 145, emphasis in the original)

Yet this is a peculiar verdict: for if, "it is agreed that *what* interpreters do is best captured *not* in mindreading terms ... then it is not at all clear why we should invoke a 'full third-person mindreading system' to explain *how* they do this" (Strijbos and de Bruin 2012, 145–146, emphasis in the original).

These debates are ongoing. As such, we are not in a position to settle cognitive ontology for social cognition here and now, and perhaps not any time in the near future either. However, the issue is not just that we do not yet know how to best characterize social cognitive tasks. As this case study reveals, it is possible that in characterizing such tasks we will need to relinquish the claim that neural structures have dedicated mindreading functions.

Moreover, the above considerations reveal why we ought to leave open the possibility that neural activity might be doing its work in social cognition by being enlisted into service, as dictated by context, to enable direct and dynamically unfolding interpersonal engagements and interactions with others.

There is a more fundamental reason to be skeptical about certain versions of the cognitive ontology enterprise – in particular any that operate with simple subtraction methods that assume, for example, that increased brain activity is a direct and reliable proxy of cognitive activity. When it comes to brains that inference is unsafe. As Anderson (2014) emphasizes, that simplistic approach to neuroimaging studies is likely inspired by an analogy that focuses on "the least appropriate aspects of the computer metaphor of the brain. In the kind of computer with which we are most familiar, activity is an indication of computation, and inactivity is a sign of rest" (140). Yet, as Anderson goes on to observe, "the brain is not like that" (140). If we accept that Anderson is correct about the basic biological facts then it follows that there is a problem for any approach to cognitive ontology that overemphasizes or trades too heavily on assumed similarities between brains of living systems and mechanical computers when trying to understand cognitive tasks.[4]

Notably, neither our analysis of the case of mindreading nor Anderson's methodological worries about dangers of over-reliance on the computer-brain metaphor are fatal to neurocentric approaches to cognitive ontology. For even if Anderson's observation is true it is open for the field to develop methods that would enable neuroscientists to make safer and more appropriate inferences about the links between brain activity and cognition. Furthermore, it is logically possible to adopt a wait-and-see stance about whether there are any dedicated neural structures for mindreading. So a neurocentric cognitive ontology is still an option but it remains to be shown that it must be in the cards or is, at least, the best bet. Adhering to neurocentricism will be attractive to anyone who assumes that there is a decidable answer to questions about what functional contribution specific regions and networks make to the performance of specific cognitive tasks. But why assume that? Why assume, as Anderson (2014) puts it, that "there *must be* an answer to such questions" (p. xix, emphasis added)? Why assume that neural structures have any stable functions at all? Why not allow that the brain is radically MacGyverish and protean – using and reusing its internal resources in inventive, on-the-fly improvisations to suit circumstance and context?

In sum, like Anderson (2014), we too "fail to see the scientific motivation for pushing onward with the project of specifying component operations that selectively engage individual neural regions" (138).[5] We have an opportunity to change direction – namely, to question whether neural regions or networks with discoverable dedicated functions are ever at play in performing cognitive tasks. It is open to us to reconsider how deeply we, and our brains, might integrate with and make use of external resources in completing cognitive tasks.

3. Anderson's alternative

Anderson (2014) proposes going another way, directing us to start in a different place. Starting from the top, his main message is: ask not what cognitive operations are implemented in individual neural regions, ask what are the psychological factors that account for the differential activity of the brain. He calls such task dependent Neuroscientifically Relevant Psychological, or NRP, factors (129).

Building on the neurocentric methods favoured by Poldrack, Halchenko, and Hanson (2009) and Klein (2012), Anderson still seeks "to measure the responses of the brain in

a multidimensional manner and try to discern the structure that underlies these responses" (Anderson 2014, 114). Anderson's approach (2014) shows it is possible to analyse the multidimensional functions of neural assemblies "using simple and coarse task categories" (139). Yet he does so while steering clear of the idea that the brain necessarily carves up into components with stable functions. Hence he offers us a quite different way of "doing justice to the underlying functional complexity of individual regions of the brain" (138).

Importantly, in Anderson's hands a multifactor analysis of individual regions of the brain is not conducted in order to discover neural structures with stable functions but to determine which areas of the brain "are active across multiple tasks" (Anderson 2014, 138). It is not single regions or a set of regions which are investigated. Instead, the activity of the whole brain is measured during a task and represented as a large multidimensional vector. The size of the vector is reduced in order to make analysis computationally tractable, for example, by using artificial neural network classifiers. This makes it possible to do cognitive neuroscience "without the analysis, decomposition, and location of component cognitive operations" (117). What we end up with are "fingerprints" (117) of the brain participating in different cognitive tasks. Those fingerprints are then functionally profiled in relation to NRP factors. The express purpose of a multidimensional investigation of the neural structures linked to NRP factors is to gain insight into the functional profile of neural assemblies across their different network states.

Thus Anderson breaks faith with those authors whom he regards as "overly-influenced by a particular computational model of brain architecture" (Anderson 2015, 74). In contrast he favours a more radical, protean approach to basic brain architecture – one that expressly embraces the idea that "an evolved (as opposed to an engineered) system will likely be characterized by the use and reuse of the same parts for many different purposes" (Anderson 2015, 74).[6]

Used in Andersonian fashion, the set of techniques for divining structure in multidimensional data mentioned above are "sensitive enough to reveal neural differences — such as diagnostic patterns of distributed neural activity — that are sufficient to establish differential responsiveness without thereby demonstrating selectivity in any region or network" (Anderson 2014, 141). Thus he identifies the central point in the philosophical head shift in re-thinking the job of neuroscience as follows: it is "not just that we do not get selectivity in the brain but that we don't need it. We can stop looking for it" (Anderson 2014, 141). His signature realization is that we can "give up on the global framing notion of component operations in a selective, computational brain and build our neuroscience instead around the idea of NRP factors differentially expressed in a dispositional brain" (Anderson 2014, 142).

This way of conceiving of the evolving network dynamics and functional interactions of a much more extensive variety is needed to help us "make better sense of the brain and what it is doing for the organism" (Anderson 2014, 208). He locates this emerging picture of the brain in the context of the E-approaches to cognition, urging that, "we seriously consider the likelihood that the best way to understand the brain's native ontology – the dimensions of the neurally-structured psychological space – is in evolutionarily-inspired, ecological, and enactive terms" (Anderson 2015, 75).

Anderson's alternative is thus aligned with and gains support from new developments in cognitive science. Research on global network dynamics and anticipatory brain dynamics, for example, suggests that brains are parts of larger self-organizing systems that are forever trying to look ahead and anticipate the future in order to ensure that they have an adequate practical grip on the world in the here and now. As such brains are not best understood as

computing incoming informational content but as rather existing in "in metastable states" that are always poised on the brink of instability (Kelso 1995, 26). Brains need to be able to "switch flexibly and quickly" between states in order to "anticipate the future, not simply react" (Kelso 1995, 26; see also Kirchhoff 2015).

Anticipatory models of the brain depict distributed brain activity as involving local connectivity across various brain areas, but where macroscopic or global influences constrain and shape "the intrinsic dynamics of thalamocortical networks and constantly create [anticipations] about forthcoming sensory events" (Engel, Fries, and Singer 2001, 704). This view of the brain is perfectly positioned to take the dynamical organization of the body and external resources as a constituting part of context-sensitive cognitive activity (Riley, Shockley, and Orden 2012). For example, Riley, Shockley, and Orden (2012) report that central pattern generators "can be reorganized when sensory and mechanical feedback trigger neurotransmitter release that functionally alters the network connectivity . . . for example, a mollusk slowly treading water changes abruptly, recruiting additional interneurons to enable rapid escape from a predator" (2012, 24).

So conceived, the contributions of the brain to the performance of cognitive tasks are always part of a larger network comprising those of the body and the environment (Gallagher et al. 2013). To see this, consider, first, that although global brain dynamics can arise from local activity, large-scale dynamics also constrains local neuronal activity. This implies that local and global neurodynamics are coupled, exhibiting a form of continuous reciprocal causation. An example of this is the slaving principle in synergetics, where microscopic processes or patterns of activity become enslaved by macroscopic or large-scale ensemble behaviour of the system in question (Deco and Jirsa 2012). Second, this kind of coupling is not intrinsic to the brain alone but exists between brain, body, and environment. Crucially, if the brain-body-environment system comprises a single and unified system, then there is no obvious way to carve this larger system into its native cognitive parts and pieces (Silberstein and Chemero 2012). And as discussed in the previous section, if Anderson (2014) is right, there is no scientifically compelling need to make such an attempt.[7]

This view of neurodynamics depicts the brain as anything but isolated from bodily activity in real-world situations. Synergies not only enslave local brain areas into coherent dynamical patterns but the dynamics break across brain and body. Understanding brains as doing fundamentally predictive work of this sort and as focused on action-oriented engagement is, arguably, perfectly in tune with the recent trends of conceiving of cognition as embodied, ecologically situated, extended, and encultured (Clark 2016, 108, 125). Going the Andersonian way asks us to abandon the view of the brain as a mechanical computer that receives and processes input and to see it, instead, as an action-oriented enabler, whose anticipatory dynamics are realized in multiple areas, over diverse temporal and spatial scales.

4.　Cognitive ontology in flux

How drastically would we need to rethink our cognitive ontology if we adopted an Andersonian alternative vision of neurodynamics? How big is the proposed change? Despite the fact that Anderson (2014) pins his colours to a "truly functionalist cognitive neuroscience" (296), he also embraces enactivism. Yet, it is not obvious that enactivism is compatible with functionalism.

Standard functionalism seeks to characterize cognition in terms of interactions between discrete component parts and the interactions between these parts are understood in terms of processes that are characterized by well-defined input and output states. By contrast, the sort

of dynamical systems approaches that enactivists promote assume an inescapably messier, loopier vision of cognition. Thus, amongst all of the E-approaches, enactivism provides perhaps the clearest test case. Enactivists conceive of cognition as a spatiotemporally extended phenomenon that "unfolds as the continuous coevolution of acting, perceiving, imagining, feeling and thinking" (Thompson 2007, 43). Their focus of interest is on the complex dynamics of how cognitive processes unfold in extensive activity that takes place across various spatiotemporal scales and spans brain, body and world, forming a dynamic singularity – a "tangle of recurrent and reentrant processes" (Colombetti 2014, 103; see also Varela 1979 and Thompson 2007).

Enactivists hold that, "in the context of this tangle . . . choosing to describe one component systems as under the control of another, receiving inputs and sending outputs back to it, amounts to taking a heteronomy or 'other-governed' perspective" (Colombetti 2014, 103). Although adopting a heteronomous perspective on cognition – namely, a perspective that sees cognition as governed by and only responsive to externally defined parameters – is clearly possible, it has real limits when approaching the cognitive activity of living systems. In particular, it has been suggested that this heteronomous perspective misses something important: it fails to "illuminate – and indeed can obscure – certain observable patterns of behavior; namely, patterns arising from the system's internal dynamics" (Thompson 2007, 50). Under a heteronomous perspective, therefore, the cognitive systems are depicted as lacking autonomy.

Those attracted to more dynamic views of cognition move away from a vision of mind as grounded in transitions between well-defined, component states and parts towards a process-based metaphysics. In contrast with the classic computational approach to cognition, dynamical systems are comprised not of states but of processes.[8] Importantly, unlike a state or event, a process is "something which goes on through time and can change as it does so" (Steward 2016, 76). Thus the focus is on questions such as "Why, for example, did the process unfold in this way rather than that? Why has it not stopped? What is sustaining it and keeping it going? What is responsible for any regular patterns we may observe in its progression?" (Steward 2016, 78). Moreover, there need not be neat and tidy answers to questions about where a process begins or ends.

The significance of the shift in cognitive science toward predictive, pragmatic and process views of cognition is perhaps best captured in a pregnant statement by Andy Clark when he says that these new ways of thinking about cognition go so far as to abandon the "last vestiges of the 'input–output' model" (2016, 139).

In this light, it is unsurprising that Di Paolo (2009) holds that "the enactive project . . . departs from any form of functionalism" (16) and Di Paolo and Thompson (2014) tell us that there is "a fundamental difference between enactive and functionalist approaches to the mind" (73).

Even so, whether enactivism is incompatible with functionalism depends on how we construe functionalism. Enactivism is incompatible with functionalism if the latter is defined in a familiar but restrictive manner such as when a functional mechanism is taken to be "a system of component parts with causal powers that are organized to perform a function" (Piccinini 2015, 1).[9]

Yet, for all that has been said, enactivism – even of the most radical kind – need not be at odds with any and every kind of functionalism – namely, it need not be at odds with functionalism simpliciter. Traditionally, functionalism has been understood in terms of computational states, but it has been proposed that functionalism may not require this commitment. For example, Piccinini (2010) claims that the most minimal functionalism only requires that a cognitive system be defined as having the functional organization

"of the brain, or any other system that is functionally equivalent to the brain" (270).[10] The minimal construal of functionalism is ultra liberal, and since – as we have seen above – questions about the nature of the brain's organization are precisely those that are up for grabs at this stage of inquiry. The important point is that if a minimal functionalism of the sort Piccinini envisages proves coherent and tenable then "functionalism may be combined with a non-computational theory of mind" (Piccinini 2010, 271). And if a minimal functionalism is compatible with non-computational theories of mind then this opens up the possibility that enactivism may be compatible with functionalism after all. However, getting clear about the precise features of minimal functionalism and deciding whether it makes, in the end, an explanatorily contribution is a matter for further investigation.

5.   Conclusion

We have not argued against the in-principle possibility of finding brain-bound cognitive mechanisms composed of neural structures with dedicated functions. Such mechanisms may, in the end, turn out to form part of our cognitive ontology. However, we have argued that in light of the current state of philosophy and science of cognition there is no reason to assume the familiar cognitivist and neurocentric framework that would make the finding of such entities inevitable. Indeed, we should be cautious about adopting such a framework precisely because it forecloses, without compelling evidence or argument, on another important metaphysical possibility.

In promoting the possibility that brains might be protean in character we are making room for thinking of cognitive tasks as performed by brains working in concert, interactively, with a range of non-neural bodily and environmental resources, rather than assuming they are always performed by neural networks or brain regions working alone.

When it comes to thinking about the metaphysics of cognitive tasks it is a live possibility that neural structures are shaped and configured in context-sensitive ways, conforming to the needs of overall tasks rather than always and everywhere depending on the activity of brain regions that divide into naturally existing, functionally stable parts and pieces. Putting all of this together, there are reasons to rethink some standard assumptions about cognitive ontology and the assumption that cognition always comes with already carved natural joints.

Notes
1.   This project is one of specifying "the entities that exist in [the cognitive] domain and the relations between them" (Lenartowicz et al. 2010, 679). The systematic descriptions of structure-function relations of neural systems are "referred to as ontologies" (Price and Friston 2005, 263).

2. In line with realism, it is assumed that "this mapping can only be successful if the cognitive constructs being mapped to the brain *are actually implemented* in the brain as separate constructs" (Lenartowicz et al. 2010, 680, our emphasis). Thus the process of articulating the correct cognitive ontology requires us to "clarify, refine, and test theories of brain and cognitive function" (Poldrack and Yarkoni 2016, 587). They are meant to "facilitate the integration of cognitive and anatomical models and organize the cognitive components of diverse tasks into a single framework" (Price and Friston 2005, 262).

3. Our proposal about the protean character of cognitive ontology fits most naturally with embodied, enactive approaches at the extreme end of the spectrum – namely, those that endorse an extensive enactivist approach to cognition (see Hutto, Kirchhoff, and Myin 2014 and Hutto and Myin 2013, Chapter 7).

4. Enactivist authors, such as Thompson (2007) and Noë (2009), stress the importance of looking at life not computers when trying to understand and model the mind. Respectively, they tell us, "life and mind share a set of basic organizational properties, and the organizational properties distinctive of mind are an enriched version of those fundamental to life. Mind is life-like and life is mind-like" (Thompson 2007, 128); and that "What biology brings into focus is the living being, but where we discern life, we have everything we need to discern mind" (Noë 2009, 41).

5. Some may be reticent about giving up on a computational, componential approach to mind because of putative past successes that the classical cognitivist framework has allegedly delivered. Take the favourite parade case of Marr's (1982) computational theory of vision as a case in point. It is frequently taken to have heralded "the arrival of visual psychology as a maturing science" (Burge 2010, 93). It is widely regarded as both "extremely impressive and influential" (Tye 2000, 83) and as one of cognitivism's most "stunning successes" (Shapiro 2014, 2). In short, it is the canonical cognitivist success story. Yet there are well-known limitations in modelling vision in Marrian terms. A major worry is that it fails to accommodate the interactive complexity of cognition because it thinks of vision in overly linear terms. Its classic style of sequential processing is deemed too slow and rigid to properly account for the dynamically up-dated, on-the-fly character of intelligent responses. Although there have been attempts to address these issues by trying to build feedback into cognitivist stories in various ways, arguably, such developments are at best epicyclical patches lacking the advantage of more radical overhauls to cognitivist thinking (see Clark 2016, 51–52 and Anderson 2014, 166–167). Thus despite having many staunch defenders – who frequently attempt to justify sticking with classical cognitivism by pointing to its seeming success – many in the field today now regard Marr's theory of vision as marred by serious explanatory limitations and empirical inadequacies.

6. In some of Anderson's earlier work a non-protean reading of reuse is present, but we are following Anderson's later work here which understands reuse more explicitly as protean. Poldrack and Yarkoni (2016) also favour these more revisionary approaches. Thanks go to an anonymous reviewer for pointing this out.

7. Of course, talk of the brain anticipating future events is likely to induce a representationalist view of anticipatory neurodynamics. The worry goes like this, or something close to it: if the brain is in the business of anticipating future events, then it does not have access to such events and must therefore represent them. But this conclusion need not follow. In the literature on dynamical systems, it is common to treat two separate pendulums as coupled, and therefore as a constituting a single nonlinear dynamical system. It is relatively straightforward to show that any dynamical system A – for example, an organism – coupled to a second dynamical system B – for example, an environment – can be understood as anticipating the dynamics of B "when it reliably covaries with the dynamics of B and it is robust to the noise inherent in the coupling" (Bruineberg and Rietveld 2014, 7). This is the notion of anticipation we have in mind when saying that the brain continuously tries to anticipate unfolding events in the immediate future. If it is not representations that maintain a connection between brain and world, what is it? Work on global network dynamics and dynamical systems theory suggests that it is *synergies*. A synergy is an assembly (typically short-lived) of processes enslaved to act as a single coherent and functional unit (Kelso 1995). Specifically, synergies "are defined as compensatory, low dimensional relations in the dynamic activities of neuromuscular components (Kelso 2009), not as static representational structures such as motor programs" (Riley, Shockley, and Orden 2012, 23).

8. In any case, talk of states when thinking about dynamical systems appears to be an idealization at best. As Spivey argues

> [Claiming] that a system was in a particular "state," X, at a particular point in time, really boils down to saying that the *average* of the system's states during that *period of time* was X. This kind of coarse averaging measurement is often a practical necessity in science, but should not be mistaken as genuine evidence for the system actually resting in a discrete stable state. (2008, 30)

9. This way of characterizing functionalism best suits the needs of classical computational theories of mind – those that deem computational mechanisms to be a special sub-class of functional mechanisms, distinguished by the fact that they "manipulate vehicles based solely on differences between portions of the vehicles in accordance with a rule that is defined over the vehicles and, possibly, certain internal states of the mechanism" (Piccinini 2015, 1). Thus, "computational states and processes are individuated functionally, i.e. formally and syntactically" (Piccinini 2015, 2).

10. This may come as no surprise. After all, as Putnam explained long ago, when first introducing functionalism, it was only ever a framework for theorizing and advancing empirical hypotheses; in itself functionalism is "the putting-forward, not of detailed scientifically 'finished' hypotheses, but of schemata for hypotheses" (Putnam 1967/1992, 54). Of course, even if enactivism is compatible with a minimal functionalism, this leaves open the question of whether a minimal functionalism adds any explanatory payoff or punch to enactivism.

Notes on contributors

Daniel D. Hutto is Professor of Philosophical Psychology at the University of Wollongong and member of the Australian Research Council College of Experts. His most recent books, include: *Wittgenstein and the End of Philosophy* (Palgrave, 2006), *Folk Psychological Narratives* (MIT, 2008). He is co-author of the award-winning *Radicalizing Enactivism* (MIT, 2013) and its forthcoming sequel, *Evolving Enactivism* (MIT, 2017). He is editor of *Narrative and Understanding Person*s (CUP, 2007) and *Narrative and Folk Psychology* (Imprint Academic, 2009). A special yearbook, *Radical Enactivism*, focusing on his philosophy of intentionality, phenomenology and narrative, was published in 2006.

Anco Peeters is a Ph.D. candidate at the University of Wollongong.

Miguel Segundo-Ortin is a Ph.D. candidate at the University of Wollongong.

ORCID

Daniel D. Hutto http://orcid.org/0000-0002-3585-4618
Anco Peeters http://orcid.org/0000-0002-2380-4154
Miguel Segundo-Ortin http://orcid.org/0000-0002-3787-0013

References

Aizawa, Ken. 2015. "What Is This Cognition That Is Supposed to Be Embodied?" *Philosophical Psychology* 28 (6): 755–775. doi:10.1080/09515089.2013.875280.

Anderson, Michael L. 2014. *After Phrenology: Neural Reuse and the Interactive Brain*. Cambridge, MA: The MIT Press.

Anderson, Michael L. 2015. "Mining the Brain for a New Taxonomy of the Mind." *Philosophy Compass,* 10: 68–77. doi:10.1111/phc3.12155.

Apperly, Ian. 2011. *Mindreaders: The Cognitive Basis of "Theory of Mind"*. New York: Psychology Press.

Bruineberg, Jelle, and Erik Rietveld. 2014. "Self-Organization, Free Energy Minimization, and Optimal Grip on a Field of Affordances." *Frontiers in Human Neuroscience* 8: 1–14. doi:10.3389/fnhum.2014.00599.

Burge, Tyler. 2010. *Origins of Objectivity*. Oxford: Oxford University Press.

Clark, Andy. 2016. *Surfing Uncertainty: Prediction, Action, and the Embodied Mind. Prediction, Action, and the Embodied Mind*. Oxford: Oxford University Press.

Colombetti, Giovanna. 2014. *The Feeling Body: Affective Science Meets the Enactive Mind*. Cambridge, MA: The MIT Press.

Deco, Gustavo, and Viktor K. Jirsa. 2012. "Ongoing Cortical Activity at Rest: Criticality, Multistability, and Ghost Attractors." *Journal of Neuroscience* 32 (10): 3366–3375. doi:10.1523/JNEUROSCI.2523-11.2012.

Di Paolo, Ezequiel. 2009. "Extended Life." *Topoi* 28 (1): 9–21. doi:10.1007/s11245-008-9042-3.

Di Paolo, Ezequiel and Evan Thompson. 2014. "The Enactive Approach." In *The Routledge Handbook of Embodied Cognition*, edited by L. Shapiro, 68–78. New York: Routledge Press.

Engel, Andreas K., Pascal Fries, and Wolf Singer. 2001. "Dynamic Predictions: Oscillations and Synchrony in Top–Down Processing." *Nature Reviews Neuroscience* 2 (10): 704–716. doi:10.1038/35094565.

Gallagher, Shaun. 2001. "The Practice of Mind. Theory, Simulation or Primary Interaction?" *Journal of Consciousness Studies* 8 (5–7): 83–108.

Gallagher, Shaun. 2008. "Direct Perception in the Intersubjective Context." *Consciousness and Cognition* 17 (2): 535–543. doi:10.1016/j.concog.2008.03.003.

Gallagher, Shaun, Daniel D. Hutto, Jan Slaby, and Jonathan Cole. 2013. "The Brain as Part of an Enactive System." *Behavioral and Brain Sciences* 36 (4): 421–422. doi:10.1017/S0140525X12002105.

Herschbach, Mitchell. 2008. "False-Belief Understanding and the Phenomenological Critics of Folk Psychology." *Journal of Consciousness Studies* 15 (12): 33–56.

Hutto, Daniel D. 2004. "The Limits of Spectatorial Folk Psychology." *Mind & Language* 19 (5): 548–573. doi:10.1111/j.0268-1064.2004.00272.x.

Hutto, Daniel D. 2008. *Folk Psychological Narratives: The Sociocultural Basis of Understanding Reasons*. Cambridge, MA: MIT Press.

Hutto, Daniel D., Michael D. Kirchhoff, and Erik Myin. 2014. "Extensive Enactivism: Why Keep It All in?" *Frontiers in Human Neuroscience* 8: 1–11. doi:10.3389/fnhum.2014.00706.

Hutto, Daniel D., and Erik Myin. 2013. *Radicalizing Enactivism: Basic Minds Without Content*. Cambridge, MA: The MIT Press.

Kelso, J. A. Scott. 1995. *Dynamic Patterns. The Self-Organization of Brain and Behavior*. Cambridge, MA: The MIT Press.

Kelso, J. A. Scott. 2009. "Synergies: Atoms of Brain and Behavior." In *Progress in Motor Control*, edited by Dagmar Sternad, 83–91. Advances in Experimental Medicine and Biology 629. New York, NY: Springer US. doi:10.1007/978-0-387-77064-2_5.

Kirchhoff, Michael D. 2015. "Experiential Fantasies, Prediction, and Enactive Minds." *Journal of Consciousness Studies* 22 (3–4): 68–92.

Klein, Colin. 2012. "Cognitive Ontology and Region-Versus Network-Oriented Analyses." *Philosophy of Science* 79 (5): 952–960. doi:10.1086/667843.

Koster-Hale, Jorie, and Rebecca Saxe. 2013. "Theory of Mind: A Neural Prediction Problem." *Neuron* 79 (5): 836–848. doi:10.1016/j.neuron.2013.08.020.

Legrand, Dorothée, and Perrine Ruby. 2009. "What Is Self-Specific? Theoretical Investigation and Critical Review of Neuroimaging Results." *Psychological Review* 116 (1): 252–282. doi:10.1037/a0014172.

Lenartowicz, Agatha, Donald J. Kalar, Eliza Congdon, and Russell A. Poldrack. 2010. "Towards an Ontology of Cognitive Control." *Topics in Cognitive Science* 2 (4): 678–692. doi:10.1111/j.1756-8765.2010.01100.x.

Mahy, Caitlin E. V., Louis J. Moses, and Jennifer H. Pfeifer. 2014. "How and Where: Theory-of-Mind in the Brain." *Developmental Cognitive Neuroscience* 9 (July): 68–81. doi:10.1016/j.dcn.2014.01.002.

Marr, David. 1982. *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*. Cambridge, MA: The MIT Press.

Noë, Alva. 2009. *Out of Our Heads: Why You Are Not Your Brain, and Other Lessons from the Biology of Consciousness*. New York: Hill and Wang.

Pessoa, Luiz. 2014. "Understanding Brain Networks and Brain Organization." *Physics of Life Reviews* 11 (3): 400–435. doi:10.1016/j.plrev.2014.03.005.

Piccinini, Gualtiero. 2010. "The Mind as Neural Software? Understanding Functionalism, Computationalism, and Computational Functionalism." *Philosophy and Phenomenological Research* 81 (2): 269–311. doi:10.1111/j.1933-1592.2010.00356.x.

Piccinini, Gualtiero. 2015. *The Nature of Computation*. New York, NY: Oxford University Press.

Poldrack, Russell A., Yaroslav Halchenko, and Stephen José Hanson. 2009. "Decoding the Large-Scale Structure of Brain Function by Classifying Mental States Across Individuals." *Psychological Science* 20 (11): 1364–1372. doi:10.1111/j.1467-9280.2009.02460.x.

Poldrack, Russell A., and Tal Yarkoni. 2016. "From Brain Maps to Cognitive Ontologies: Informatics and the Search for Mental Structure." *Annual Review of Psychology* 67 (1): 587–612. doi:10.1146/annurev-psych-122414-033729.

Price, Cathy J., and Karl J. Friston. 2005. "Functional Ontologies for Cognition: The Systematic Definition of Structure and Function." *Cognitive Neuropsychology* 22 (3–4): 262–275. doi:10.1080/02643290442000095.

Putnam, Hilary. 1967/1992. "The Nature of Mental States." In *The Philosophy of Mind*, edited by Brian Beakley and Peter Ludlow, 51–58. Cambridge, MA: The MIT Press.

Ratcliffe, Matthew. 2009. "There Are No Folk Psychological Narratives." *Journal of Consciousness Studies* 16 (6–1): 379–406.

Riley, Michael A., Kevin Shockley, and Guy Van Orden. 2012. "Learning from the Body About the Mind." *Topics in Cognitive Science* 4 (1): 21–34. doi:10.1111/j.1756-8765.2011.01163.x.

Saxe, Rebecca, and Anna Wexler. 2005. "Making Sense of Another Mind: The Role of the Right Temporo-Parietal Junction." *Neuropsychologia* 43 (10): 1391–1399. doi:10.1016/j.neuropsychologia.2005.02.013.

Shapiro, Lawrence A. 2014. "Radicalizing Enactivism: Basic Minds Without Content, by Daniel D. Hutto and Erik Myin." *Mind* 123 (489): 213–220. doi:10.1093/mind/fzu033.

Silberstein, Michael, and Anthony Chemero. 2012. "Complexity and Extended Phenomenological-Cognitive Systems." *Topics in Cognitive Science* 4 (1): 35–50. doi:10.1111/j.1756-8765.2011.01168.x.

Spaulding, Shannon. 2010. "Embodied Cognition and Mindreading." *Mind & Language* 25 (1): 119–140. doi:10.1111/j.1468-0017.2009.01383.x.

Spivey, Michael. 2008. *The Continuity of Mind*. Oxford Psychology Series 44. Oxford: Oxford University Press.

Steward, Helen. 2016. "Making the Agent Reappear: How Processes Might Help." In *Time and the Philosophy of Action*, edited by Roman Altshuler and Michael J. Sigrist, 67–83. Routledge Studies in Contemporary Philosophy 83. New York: Routledge.

Strijbos, Derek W., and Leon C. de Bruin. 2012. "Making Folk Psychology Explicit. The Relevance of Robert Brandom"s Philosophy for the Debate on Social Cognition." *Philosophia* 40 (1): 139–163. doi:10.1007/s11406-010-9288-z.

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

Tye, Michael. 2000. *Consciousness, Color, and Content*. Representation and Mind. Cambridge, MA: The MIT Press.

Varela, Francisco J. 1979. *Principles of Biological Autonomy*. New York: North Holland.

Zahavi, Dan. 2011. "Empathy and Direct Social Perception: A Phenomenological Proposal." *Review of Philosophy and Psychology* 2 (3): 541–558. doi:10.1007/s13164-011-0070-3.