Mathematical Cognition

A Case of Enculturation

Richard Menary

Most thinking about cognition proceeds on the assumption that we are born with our primary cognitive faculties intact and they simply need to mature, or be fine-tuned by learning mechanisms. Alternatively, a growing number of thinkers are aligning themselves to the view that a process of enculturation transforms our basic biological faculties. What evidence is there for this process of enculturation? A long period of development, learning-driven plasticity, and a cultural environment suffused with practices, symbols, and complex social interactions all speak in its favour. In this paper I will sketch in outline the commitments of the enculturated approach and then look at the case of mathematical cognition as a central example of enculturation. I will then defend the account against several objections.

Keywords
4E cognition | Ancient number system | Arithmetical cognition | Cognitive integration | Cultural inheritance | Discrete number system | Enculturation | Evolution of cognition | Evolutionary continuity | Mathematical cognition | Niche construction | Symbol systems | Symbolic thought

Author

Richard Menary
richard.menary@mq.edu.au
Macquarie University
Sydney, NSW, Australia

Commentator

Regina Fabry
fabry@students.uni-mainz.de
Johannes Gutenberg-Universität
Mainz, Germany

Editors

Thomas Metzinger
metzinger@uni-mainz.de
Johannes Gutenberg-Universität
Mainz, Germany

Jennifer M. Windt
jennifer.windt@monash.edu
Monash University
Melbourne, Australia

1 Introduction

Since cognitive science took an ecological turn it has been casting around for new frameworks in which to conduct its main business: experimental research. Those who have taken the ecological turn are convinced that classical and brain-bound frameworks don't provide the necessary conceptual and experimental tools required to make sense of cognition in the wild (Hutchins 1995). A number of alternative frameworks have been proposed, with embodied cognition the most frequently adopted. The theoretical framework one uses to understand cognition has profound empirical consequences for scientific practice. For example, it influences what we consider to be the relevant phenomena of interest, what questions we ask about them, how we design and perform experiments, and how we interpret results (Beer 2000). The theoretical framework of classical computation, for example, approaches cognitive processing as a matter of input represented symbolically, which is then syntactically processed according to stored knowledge that the system has. It proposes a single "sandwich style" layer of cognitive processing, involving input, computation, and output (Hurley 2010).

The theoretical framework of CI (cognitive integration; Menary 2007) proposes something

Menary, R. (2015). Mathematical Cognition - A Case of Enculturation.
In T. Metzinger & J. M. Windt (Eds). *Open MIND:* 25(T). Frankfurt am Main: MIND Group. doi: 10.15502/9783958570818

altogether different: multiple cognitive layers where neural, bodily, and environmental processes all conspire to complete cognitive tasks. Although the framework is unified by a dynamical systems description of the evolution of processing in the hybrid and multi-layered system, it recognises the novel contributions of the distinct processing profiles of the brain, body, and environment. Furthermore, the CI framework explains our cognitive capabilities for abstract symbolic thought by giving an evolutionary and developmental case for the plasticity of the brain in redeploying older neural circuits to new, culturally specific functions—such as reading, writing, and mathematics (Menary 2014). I call this a process of enculturation.

This paper seeks to outline the phylogenetic and ontogenetic conditions for the process of enculturation. It will take mathematical cognition, particularly the evolutionary basis for mathematical cognition, as a core example of enculturation. In so doing, I hope to have given an account of why enculturation exists, how it happens, and in what ways it can be defended against objections. In the first section I will explore the relationship of CI to cognition embodied, embedded, enacted, extended (4E) cognition and then explain why social and cultural practices are important to the process of enculturation. In the second section I will outline the core concepts required to make sense of enculturation: continuity, transformation, novelty, and uniqueness. The third section will introduce the example of mathematical cognition, moving from the evolutionary basis for numerosity and numerical cognition to the precise operations of mathematics. The fourth section will give an account of mathematical cognition as a case of enculturation. In the final section I outline two possible objections and respond to them.

2 Where does CI sit in the 4E landscape?

Traversing the 4E landscape one rises from the lowlands of weakly embodied and embedded cognitive science to the giddy heights of strong embodiment and embedding. Embodied cognition is the thesis that at least some of our cognitive states and processes are constituted by bodily processes that are not brain-bound. Embedded cognition is the thesis that our cognitive systems are located in and interact with the surrounding physical and social environment. Enactive and extended approaches to cognition inhabit the rarefied atmosphere of the strongly embodied and embedded peaks. However, there are important differences between enaction and extension and between those variants and CI. To determine where CI and enculturation sit in the 4E landscape, I will use a dimensional analysis I first introduced in Menary (2010).

> Embodied mind
> **Embodied mind weak**: the mind/brain is embodied (compatible with internalism/individualism Smart 1959; Stich 1983)
> **Embodied mind moderate**: some of our mental and cognitive processes and states depend[1] upon our non-neural body (Gallagher 2005; Gallese 2008)
> **Embodied mind strong**: some of our mental and cognitive processes and states are constituted by processes of the body acting in and on the environment (compatible with enactivism Varela et al. 1991, and CI Menary 2007)

> Embedded mind
> **Embedded mind weak**: All the perceptual inputs to and behavioural outputs from cognitive systems are found in the environment (compatible with internalism/individualism Adams & Aizawa 2008; Rupert 2009)
> **Embedded mind moderate**: Mental and cognitive states and processes are scaffolded or causally depend upon the environment (Sterelny 2003; Wheeler 2005)
> **Embedded mind strong**: Some mental and cognitive processes and states are integrated with environmental states and processes into a single system (compatible with extended mind Clark 2008, this collection; Menary 2007; Rowlands 2010)

1 Here we might take dependence simply to be a causal, and not a constitutive, relation. Perhaps my gesturing in a particular way causes my recalling a word.

Weakly embodied mind is just the old thesis that the mind is identical to the brain. One can be an individualist and hold to this form of embodiment, and I won't consider the implications of the view here. The work of some[2] embodied cognition researchers will fall under the moderate sense of embodiment. For example, those who attempt to show that concepts or word-meanings are causally dependent upon sensorimotor areas of the brain (Glenberg 2010; Gallese 2008) commit to a moderate sense of embodiment. The strong sense of embodiment focuses on how cognition is constituted by bodily interaction with the environment, and I shall focus on the discussion here. CI and enactivism occupy this region of the environment, but with different emphases on the nature of the interaction and the evolutionary continuity of simple and complex cognitive systems. CI also occupies the strongly-embedded region, but I shall deal with the relation between CI and cognitive extension in the next sub-section.

Enactivism (excluding its radical variant)[3] allows that even simple living systems are cognitive. Enactivists are committed to the continuity of life and mind and so they propose cognitive and even mental states and processes[4] for much simpler biological systems than would CI (Varela et al. 1991).[5] Whilst I am sympathetic with the commitment to continuity between simple cognitive systems and complex cognitive systems, it is questionable whether we should argue that simply being a living organism provides sufficient cognitive complexity for conscious experience and sense (or meaning) making.

CI does not require us to think that complex cognitive and mental phenomena, such as conscious experience, are shared by all living organisms whatever their complexity or simplicity. This is to assume that the properties of complex cognitive systems will be found even in very simple cognitive systems. According to CI, this gets things the wrong way round: there is a continuity from very simple systems that interact with their environments, by having mechanisms that track or detect salient features of their environments, to complex systems that have a wider range of cognitive capabilities (traits) including memory, inference, communication, problem solving, social cognition, and so on. By contrast a phylogeny of cognitive traits would show the distribution of those traits (across species) and help us to understand both the evolutionary pressures that produce more complex kinds of cognitive systems and the innovations that bring about new traits.[6]

CI provides a phylogenetic and ontogenetic basis for when bodily interactions are cognitive processes. Along with niche constructionists (Laland et al. 2000), CI maintains a phylogeny of hominid cognition in terms their active embodiment in a socially constructed cognitive niche. Ontogentically, neonates acquire cognitive abilities to create, maintain, and manipulate the shared cognitive niche, including tools, practices, and representational systems. Cognitive processing often involves these online bodily manipulations of the cognitive niche, sometimes as individuals and sometimes in collaboration with others. CI has a unique position on the 4E landscape, because it is the first framework to propose that the co-ordination dynamics of integrated cognitive systems are jointly orchestrated by biological and cultural functions. What, though, are the cultural functions in question?

2.1 Cognitive practices as cultural practices

Both CI and extended mind (EM) occupy the strong embedding region, but they do so in different ways. Here I will differentiate CI as a thesis of enculturation from Clark's organism-

2 One could look at a classic paper on mind/brain identity such as Smart (1959).

3 See Thompson (2007) for an account of the life-mind continuity, Stewart et al. (2010) for a volume dedicated to enactivism, and Hutto & Myin (2013) for a self-proclaimed radical variant.

4 See for example Barbaras (2010), which argues that to live is to have intentional consciousness of living.

5 Interestingly, radical enactivsts appear to agree with CI on this issue; see Hutto & Myin (2013, p. 35). However, the radicals have a problem bridging the gap between basic cognitive processes and enculturated ones, since they think that meaning, or content, can only be present in a cognitive system when language and cultural scaffolding is present (Hutto & Myin 2013). That, of course, doesn't sit well with evolutionary continuity.

6 See for example Sterelny's cognitive phylogeny in Sterelny (2003) and Godfrey-Smith's complexity thesis in Godfrey-Smith (1996). See MacLean et al. (2012) for an overview of the problems for a comparative phylogeny.

centred approach to EM. Cognitive integration is a model of how our minds become enculturated. Enculturation rests in the acquisition of cultural practices that are cognitive in nature. The practices transform our existing biological capacities, allowing us to complete cognitive tasks, in ways that our unenculturated brains and bodies will not allow. Cultural practices are patterns of action spread out across cultural groups (Roepstorff 2010; Hutchins 2011; Menary 2007, 2010, 2012). Cognitive practices[7] are enacted by creating and manipulating informational structures[8] in public space. This can be by creating shared linguistic content and developing it through dialogue, inference, and narrative; or it can be by bodily creating and manipulating environmental structures, which might be tools or public and shared representations (or a combination of both). Examples of linguistically mediated action include self-correction by use of spoken (or written) instructions, co-ordinating actions among a group, or solving a problem in a group by means of linguistic interaction. Examples of creating and manipulating public and shared representations include using a graph to represent quantitative relationships; using a diagram to represent the layout of a circuit or building; using a list to remember a sequence of actions; or to solve an equation, to mathematically model a domain, to make logical or causal connections between ideas, and so on. Practices can be combined into complex sequences of actions where the physical manipulation of tools is guided by spoken instructions, which are updated across group members. A simple example of a group brainstorming with one member writing out the answers would be an example of a complex of collaborative cognitive practices.[9] Cognitive practices are culturally endowed (bodily) manipulations of informational structures.

Practices govern how we deploy tools, writing systems, number systems, and other kinds of representational systems to complete cognitive tasks. These are not simply static vehicles that have contents; they are active components embedded in dynamical patterns of cultural practice. Practices are public, and they are also embodied and enacted.[10] We embody practices: they become the ways in which we act, think, and live. They structure our lifeways (although not exclusively).

CI does not deny that much thinking takes place offline in the brain, but it does take the online and interactive mode of thought to be adaptive. Again, this line of thought has precursors,[11] but CI, uniquely, takes interactive thought as a basic category,[12] which is then scaffolded by culturally evolved practices. Practices stabilise and govern interactive thought across a population of similar phenotypes. The stable patterns of action can then be inherited by the next generation, because the practices have become settled and are part of the developmental niche in which the minds of the next generation grow. Our brains co-adapted to the stable spread of practice and its role in ontogeny—resulting in the slow evolution of the cultural brain.

The focus upon practice and culture marks cognitive integration out from variants of extended cognition, such as Clark's organism-centred approach to extension (2008). Clark's organism centred approach takes the assembly of extended cognitive systems to be controlled by the discrete organism, and brain, at the centre of it. He thereby reduces the role of cultural practices in large or small groups of organ-

7 I don't mean to suggest that there can't be other effects of cognitive practices, but since practices are just the cultural formalisation of patterns of action across a population, or group, cognitive practices are tied directly to these patterns of action. I can't provide a detailed origin account for cognitive practices here, but see Menary (2007, Ch. 5) for an early attempt to do so. However, the account of mathematical cognition I give in the next two sections provides an example of how such an account would be likely to look.

8 The primary cases I am thinking of are public systems of representation, including spoken language. However, I don't want to rule out cases involving tools, bodily gestures, artistic or bodily adornments, and the intelligent use of space and objects.

9 For two very good overviews of collective or group cognition see Theiner (2013) and Huebner (2013).

10 Jennifer Windt helpfully pointed out that practices can be thought of as public, because they are embodied and enacted. I think that this is just right: practices are patterns of action spread across a population. However, I am inclined to think that practices are not simply reducible to the bodily actions of individuals. Whilst doing long multiplication requires a bodily action of me, what I am doing cannot be described exclusively in terms of those bodily actions. The practice is a population, or group level phenomenon, not an individual one.

11 The classical pragmatists, particularly Peirce and Dewey, held that thought was interactive. See Menary (2011) for a description of pragmatist approaches to thought, experience and the self.

12 See Menary (2007, Ch. 5), where I make a detailed evolutionary case.

isms in the explanation of cognitive assembly. "Brains are special, and to assert this need mark no slippery-slope concession to good old-fashioned internalism as an account of mind. It is fully consistent with thinking (as I do) that Hutchins is absolutely right to stress the major role of transmitted cultural practices in setting the scene for various neurally-based processes of cognitive assembly" (Clark 2011, p. 458). On Clark's view, cultural practices only set the scene for the real work of integration to be done by the brain. Whilst it is arguable whether Clark's position is a return to "good old fashioned internalism," he certainly does not give cultural practices a central role in assembling and orchestrating cognitive systems.[13] Hutchins, by contrast, is committed to a full-blooded enculturated approach:

> [t]he ecological assemblies of human cognition make pervasive use of cultural products. They are always initially, and often subsequently, assembled on the spot in ongoing cultural practices. (2011, p. 445)

CI is the only variant of strong embedding (including EM) to explain the role of cultural practices in assembling integrated cognitive systems. Cognitive practices are inherited as part of the developmental niche and have profound transformative effects on our cognitive abilities. This leads us to the main concepts required to understand these transformations as a process of enculturation.

3 Enculturation: The main concepts

In this section I define and explain the main concepts required to understand enculturation, other than the already explored concepts of integration and practice. I will develop the concepts of evolutionary continuity, behavioural and neural plasticity, transformation and innovation, or novelty and uniqueness. In particular I will emphasise the phylogenetic and ontogenetic bases for modern human cognitive capacities.

3.1 Evolutionary continuity

The concept of evolutionary continuity results from the fact that evolution occurs gradually with complex structures evolving over many generations. Over long periods of time these gradual changes accumulate, resulting in large differences. Consequently, changes to a phenotype occur in slow cumulative steps over long periods of time and do not appear in a single mutational step. Evolutionary continuity demands that modern human minds evolved from earlier archaic variants. Doubtless modern minds differ from archaic minds in important respects, but these differences must have evolved over long periods of time, through slow cumulative mutational changes to the genotype. Even so, we should expect some of our archaic traits to remain, and for more modern variants to be built on top of them. One obvious example of this is the evolution of the human brain.

The evolution of the human brain can, to some extent, be seen in the gradual increase of cranial capacity, but some of the most important changes have been in the reorganisation of cortical circuitry and interconnectivity (Hoffman 2014). Although the evolution of the human brain can be understood in terms of increasing encephalization and increased connectivity between brain regions, the human brain has essentially the same set of structures as any other primate brain.[14] Modern brains evolved from archaic brains and share the same evolutionary constraints as other primates: "the similarity in brain design among primates, including humans, indicates that brain systems among related species are internally constrained and that the primate brain could only evolve within the context of a limited number of potential forms" (Hoffman 2014, p. 5). Modern minds are still partly archaic.

13 If this is an accurate portrayal of Clark's position (and I have tried to carefully use his own words) then, despite his protestations to the contrary, it appears to be a return to internalism, at least for the most central and important cognitive processes. If the brain carrys out all the important cognitive operations, then Clark's position would be a moderate embedded cognition for core cognitive abilities and an extended approach only to some of the more peripheral cases.

14 "Although species vary in the number of cortical areas they posses, and in the patterns of connections within and between areas, the structural organization of the primate neocortex is remarkably similar" (Hoffman 2014, p. 4).

It is important to think of evolutionary continuity as running from archaic to modern. We should try to avoid anthropomorphic tendencies to project modern cognitive capacities backwards into the hominin lineage or across to primate species. For example, humans are excellent social cognisers, but it does not follow from this that we should expect other primates to have a theory of mind.[15] The evolutionary pressures under which humans evolved and the capacities for complex social cognition might have been very different from those under which other primates evolved. Consequently, we should be searching for archaic precursors to modern cognitive capacities. For example, we might expect that given the increasing social pressures in hominid social groups there would be precursors to modern social cognition and that these precursors would have been adaptive solutions (Shultz et al. 2012). Modern human social cognition would then be an evolutionary consequence of increasing variation in the complexity of social organisation and interaction (Sterelny 2003).

I am committed to another sense of continuity: that between biology and culture. Culture is not, as a category, distinct from the biological. Although culture is sometimes thought of as floating free of our biological nature and sometimes as being highly constrained by it, I shall assume that genes and culture co-evolve[16] mutually, influencing and constraining one another. Therefore I shall accept no culture–biology dualism in this paper. Indeed I shall adopt a cultural inheritance model of cognitive evolution (of the niche construction kind). However, I shall always do so with archaic origins in mind. Archaic origins matter to cognitive evolution and they matter to the way our brains develop during the lifespan.[17]

15 Indeed, it is questionable whether humans deploy a theory of mind, or at least, perhaps they only do so on rare occasions (Hutto 2008; Andrews 2012). Andrews has also argued that we may share a number of "mind reading" strategies with other primates that don't involve theory of mind (2012).

16 See below for a niche construction account of gene-culture co-evolution. I favour such an account because it helps us to understand how a developmental niche could have cumulative downstream evolutionary effects on phenotypes (Sterelny 2003).

17 They matter because they are part of the developmental biases that produce a robust phenotype.

In the "modern synthesis" there is only one line of inheritance, and that is genetic inheritance. More recently, biologists (Odling-Smee et al. 2003) have proposed that there are other lines of inheritance: ecological inheritance and cultural inheritance (Boyd & Richerson 2005). Many organisms construct the niche in which they live, mate, hunt, and die. Niche constructors modify the ancestral environment, and these modifications are bequeathed to the next generation. Modifications encompass physical alterations, such as living in mounds or constructing hives, as well as cultural artefacts, practices, and institutions. Over long periods these alterations to the niche can have profound effects on the phenotype. For example, the ubiquitous niche constructions of termites, burrows and mounds, have profoundly altered their morphology and behaviour (Turner 2000).

Humans are also ubiquitous niche-constructors. They physically alter their environment and they also epistemically, socially, and culturally engineer the environment (Sterelny 2003, 2010; Menary 2007). Humans are born into a highly structured cognitive niche that contains not only physical artefact, but also representational systems that embody knowledge (writing systems, number systems, etc.); skills and methods for training and teaching new skills (Menary & Kirchhoff 2014); and practices for manipulating tools and representations. Inherited cultural capital is a real and stable feature of the socio-cultural environment, including a great variety of knowledge systems, skills, and practices across a variety of domains of human action. As such, human cultural niches provide neonates with rich developmental niches. It is in these developmental niches that humans acquire cognitive practices.

Cognitive practices are products of cultural evolution, evolving over faster timescales than biological evolution. Writing systems, for example, are only thousands of years old; consequently, it is highly unlikely that there is a "reading gene" or even an innate specialised "reading module." This is important: cognitive capacities for reading and writing, mathematics, and other culturally recent forms of cognition could not be biological adaptations (that evolved over long periods of time). The timescales for their evolution are too short. It follows that the capacity for culturally recent forms of cognition must be acquired through learning and training.

Although there are no innate specialized modules for these recent forms of cognition, cortical circuits with which we are endowed through evolution are transformed to perform new culturally recent cognitive functions, even though they evolved to perform different functions. Recent cognitive innovations aside, there are good reasons to expect that evolution has driven us to think by interacting with the environment and that this is adaptive (Sterelny 2003 2012; Menary 2007; Wheeler & Clark 2008). However, it is the scaffolding of cultural practices that orchestrates the interactions—as in the case of written language and mathematics.

Structured socio-cultural niches have had profound evolutionary consequences in the hominin lineage. Structured niches have co-evolved with human phenotypic and developmental plasticity. We have evolved to be a behaviourally plastic species (Sterelny 2012) as well as a cultural species. In this co-evolution we have developed all manner of skills, practices, and activities. Why, though, are we so peculiarly behaviourally plastic? One good answer to this question is that human behavioural and developmental plasticity is an adaptive response to the variability and contingency of the local environment (Finlayson 2009; Sterelny 2003, 2012; Davies 2012). This is an alternative to the view that we are adapted to a pleistocene hunting and gathering environment—a view relied upon by many evolutionary psychologists (Barkow et al. 1992).

Critical to a co-evolutionary account of cultural practices is the evolution of human plasticity. Given that there is such a variety of cultural activity, we need an account of human evolution that will allow for variability in human behaviour. Second, we need a model that explains how innovations in our cultural niche are inherited and propagated, leading to changes in behaviour over time. The niche construction model explains how both of these causal factors could come into play. In the subsections below, I outline the importance of behavioural and neural plasticity, the concept of transformation, and those of novelty and uniqueness.

3.2 Behavioural and neural plasticity

In evolutionary terms, humans are capable of developing a wide range of skills that allow them to cope with a wide variety of environments (and their contingencies). For example, even where skills are (broadly) of the same type, such as hunting, they will vary in how they cope with the differences in local environments—think of the differences in environments between Aboriginal hunters in the Pilbara desert, hunter-gatherers in the Central American rainforests, and Inuit seal-hunters (Sterelny 2003, p. 167).

Development is extended in modern humans relative to other species. Humans take a long time to learn how to walk and talk, and much, much longer to develop fine-grained manual and cognitive skills such as reading and writing. Other primates have much faster developmental timescales. While this might make humans more dependent on their caregivers for longer, it also allows them to refine skills and acquire a greater array of them before entering adulthood.

Through cultural inheritance, knowledge, skills, and artefacts are passed on to the next generation, but learning environments and learning techniques are also passed on so that the next generation can acquire and be transformed by the inherited cultural capital. This last point is important for our purposes, because developmentally plastic humans need scaffolded learning environments in which to develop.[18]

How, though, are we capable of acquiring these new cultural capacities in development? Through neural plasticity. Rather than the process of synaptogenesis or lesion-induced plasticity,[19] the kind of plasticity I will discuss here is

18 If the cognitive abilities for manipulating artefacts and representations are not innate, then a scaffolded learning environment helps to explain how we acquire them.

19 Many neurological studies of plasticity focus on synaptogenesis, the florid growth of grey matter and then the consequent pruning, or the

what I call learning driven plasticity (see Menary 2014). Learning driven plasticity (LDP) can result in both structural and functional changes in the brain. Structurally, LDP can result in new connections between existing cortical circuits. Functionally, LDP can result in new representational capacities (the ability to represent public symbolic representations such as alphabets and numerals) and new cognitive abilities, such as mathematics,[20] reading, and writing (Dehaene 2009; Ansari 2012). It should come as no surprise that learning drives structural and functional changes in the brain, given the extended developmental period in humans and the late development of the cortex (Thatcher 1991). The brain changes, not just because of maturation, but also because of learning:

> [w]hen children learn to read, they return from school 'literally changed'. Their brains will never be the same again. (Dehaene 2009, p. 210)

Famously, Dehaene argues that a region of the occipito-temporal junction (which he calls the VWFA, visual word form area) that is part of a wider network for recognising faces, objects, and even abstract shapes (such as chequer patterns), alters its function to recognise written symbols in alphabets and even logographic scripts such as kanji (Dehaene 2009). This is due to the plasticity of that area of the brain, where the functional shift is due to scaffolded learning.[21] "Scanning of 'ex-illiterate' adults who learned to read during adulthood has demonstrated that the VWFA is highly plastic, even in adults, and quickly enhances its response to letter strings as soon as the rudiments of reading are in place" (Dehaene & Cohen 2011, p. 259). Even those who are not convinced that a specialised region for "word recognition" is acquired once we learn to read admit that the occipito-temporal junction is part of a reading and writing circuit (e.g., Price & Devlin 2011).

We have evolved to be phenotypically and developmentally plastic. This is in no small part due to the plasticity of our brains. Our developmentally plastic brains exhibit learning-driven plasticity. When the brain is coupled to a highly scaffolded learning environment it is profoundly transformed, structurally and functionally, and consequently we are cognitively transformed in the profoundest way.

3.3 Transformation

The transformation thesis can be given a simple formulation: cognitive transformations occur when the development of the cognitive capacities of an individual are sculpted by the cultural and social niche of that individual. Cognitive transformations result from our evolved plasticity and scaffolded learning in the developmental niche. In the previous sub-sections an account was given of the effects of cultural inheritance and niche construction on hominid evolution. The result is phenotypic plasticity, and in the cognitive case the co-evolution of neural plasticity and scaffolded learning. However, the point of the transformation thesis is to drill down into the process of acquiring knowledge, skills, and cognitive abilities via learning-driven plasticity and scaffolded learning. It does this by showing how transformations are a result of the role of cognitive practices in development. Practices structure the niche; they transform plastic brains via learning driven plasticity and result in new cognitive abilities.

During the learning and training of a skill, such as flaking an arrowhead, or a shot in tennis or cricket, we are guided by the norms for the correct actions that make up the skilled practice. A parallel case can be made for cognitive abilities such as mathematics. The neophyte mathematician gains mastery over the cognitive norms[22] by which numerals, operators, and other symbols are created and manipulated. Vygotsky expresses this in the claim that children, "master the rules in accordance with which ex-

synaptic death of many of those neurons in the so-called critical period of childhood. There are a large number of studies of neural damage, often by stroke or injury, where cortical circuitry becomes damaged and its function impaired, but where other areas of the cortex can take on the impaired function. (See Huttenlocher 2002 for an overview.)

20 I will be defending an account of mathematical cognition in section 4.

21 See Menary (2014) for a discussion of plasticity and the VWFA.

22 For an account of cognitive norms see Menary (2007), Chapter 6.

ternal signs must be used" (Vygotsky 1981, pp. 184–185). Initially the child masters the creation and deployment of spoken linguistic signs (and later written signs) through the scaffolding of parents and caregivers. However, this process is not simply a matter of gaining new representations; it is also one of gaining new abilities.

Neophytes go through a process of dual-component transformation: they learn how to understand and deploy public symbolic representations and they learn how to create and manipulate inscriptions of those symbols in public space (Menary 2010). In so doing, they learn mathematical and linguistic concepts and they learn how to manipulate inscriptions to complete cognitive tasks. When learning the manipulative techniques, the first transformation is one of the sensory-motor abilities for creating and manipulating inscriptions: we learn algorithms like the partial products algorithm[23] and this is an example of the application of a cognitive practice. This is something we learn to do on the page and in the context of a learning environment, in public space, before we do it in our heads. Our capacities to think have been transformed, but in this instance they are capacities to manipulate inscriptions in public space. This is a way of showing that the transformation of our cognitive capacities has recognisably public features. This ought not to be a surprise, given that the cognitive niche is socially and culturally constructed and is structured by socio-cultural practices. Symbol systems, such as those for written language and mathematics, are not impermanent scaffolds that we shrug off in adulthood, but are permanent scaffolds that indelibly alter the architecture of cognition.[24]

The transformatory position is quite different from that held by Clark or Sterelny. In particular it holds that our basic cognitive capabilities are transformed in development and that the dual component transformation results in a distinct functional redeployment of neural circuitry and new abilities to bodily manipulate structures in public space. Cognitive tasks can be completed by manipulating written symbols in public space or by off-line strategies for completing algorithms, or a combination of both. This conclusion sits happily with the idea that thought is interactive and governed by practices.

The main difference between the position outlined here and Clark's (e.g., 2008), is that Clark does not explain cognitive extension in terms of the transformation of basic cognitive resources during development in a socio-cultural niche (although he does acknowledge the importance of symbolically structured niches). Rather, he thinks that basic biological resources are not really transformed but simply dovetail to external symbols (Clark 2008, 2011). Sterelny (2010) concentrates on cognitive scaffolding, but does not think that the manipulation of symbols in public space is constitutive of cognitive processing. The enculturated approach of CI answers questions that are problematic for both Clark and Sterelny:

1. How do we learn to complete cognitive tasks that require the manipulation of symbols in public space?
2. Assuming that cognitive processing criss-crosses between neural space and public space, how does it do this?

The first question is hard for Clark since he does not think that our basic cognitive resources get transformed, at least in the way that I have presented here. The second question is hard for Sterelny because he limits himself to a scaffolded view of cognition rather than an extended view. Consequently, manipulations of symbols in public space are not cognitive processes for Sterelny.[25]

CI as a process of enculturation requires a robust transformation thesis. A robust transformation thesis is warranted by phenotypic and neural plasticity, in particular by learning driven plasticity. Novel and unique public systems of representation drive the transformation of our existing cognitive abilities.

3.4 Novelty and uniqueness

Sometimes symbols and tools provide us with novel functions: they radically extend our cap-

23 I'll look at this example in detail in section 5.
24 I take this issue up again in section 4.1.
25 Or they might be assuming that Sterelny does not care either way; in private communication Sterelny indicated that he does not think that boundary disputes are of much interest.

abilities in some sphere. Take the humble hand axe. Very crude hand tools have been discovered dating as far back as 2.6 mya (million years ago; Toth & Schick 2006), since then there has been evidence of a hominid capacity for cumulative cultural inheritance "which was ultimately to transform Homo sapiens into the richly cultural species we are today" (Whiten et al. 2011). However, the capacity for developing novel functions and transmitting them to the next generation with high fidelity appears to be a more recent innovation, as evidenced by the long periods of relative stability in technological development in the early hominids and archaic humans. It also appears to be an innovation unique to the homonin lineage (Whiten et al. 2011). The Oldowan period begins in the lower paleoloithic with Homo Habilis around 2.6 mya, being taken up by Homo Erectus and Ergaster and ending at about 1.8 mya (Lycett & Gowlett 2008). The tool types and process of manufacture remain consistent during this period, with some refinement and novelty (Lycett & Gowlett 2008), where the main tool types were choppers and scrapers or mode 1 tools (Semaw et al. 2003).

Homo Habilis is unique in that it is the first hominid to make tools that were made to endure and be re-usable (it is likely that earlier anthropocines used naturally-occurring objects as tools that were disposable; Jeffares 2010).

> Oldowan toolmaking involves the production of sharp-edged flakes by striking one stone (the core) with another (the hammerstone). Effective flake detachment minimally requires visuomotor coordination and evaluation of core morphology (e.g., angles, surfaces) so that forceful blows may reliably be directed to appropriate targets (Stout et al. 2008, p. 1940).

There is a clear transition to Achulean technology at around 1.7 mya with the appearance of Erectus/Ergaster. The main innovation for Achulean technology was the bifacial handaxe—a handheld cutting tool with two cutting sides. The real explosion in novelty occurs in the upper paleoloithic period, from 50,000 years ago (ya) to 10,000 ya (or to just before the advent of agriculture and the neolithic period), with genuine novelty in tool production and use and cultural diversification. In this period we begin to see evidence of art, including paintings and sculpture, fishing, jewellery, burial, evidence of musical activity, and all the hallmarks of behaviourally modern humans. It is in this period that the combination of inherited cultural capital, with phenotypic and learning-driven plasticity, complex social relations and language results in an explosion of cultural and behavioural diversity.

It is also in this period that we begin to find evidence of proto-numerical and writing systems as novel representational innovations. Simple tally notch systems on bone fragments have been dated to between 35,000 and 20,000 ya, and may have been used for a variety of purposes, the most obvious being to keep track of economic exchanges. However, it is far easier and more economical to keep track of larger amounts using a single symbol, rather than a one-to-one correspondence of marks with things.

The complex social and economic pressures that required tracking exchanges involving increasingly large numbers would be the kind of socio-economic pressures that produced symbolisation of quantity. Social and cultural pressures can drive evolutionary novelty, in this case symbolisation and uniqueness—symbolic representations are unique in both type and property, no other animal produces written symbols to represent concepts. Symbols have unique properties that allow for operations—addition, subtraction, multiplication, division, and so on that are much harder (if not unlikely) without them.

Early symbolic number systems date from between 3000–4000 BCE, but genuinely abstract symbol systems are even more recent—about 1000–2000 BCE. The invention of symbol systems is too recent to be a genetic endowment, but is inherited as cultural capital and acquired through high-fidelity social learning (which is in turn dependent upon neural plasticity).

The phylogeny of hominid tool-use is one of hard-won innovation and retention. Modern humans have developed high-fidelity modes of transmitting cultural capital vertically and horizontally. The socio-cultural pressures that led to humans innovating symbolic representational systems are unique and very recent. Fortunately, modern human minds are flexible enough to both innovate and reliably acquire those innovations in ontogeny.[26] This flexibility makes modern human minds unique, and in the case of mathematical cognition unique amongst all our primate relatives.

The next section outlines mathematical cognition as a case of enculturation, and there I will explore the example of mathematical cognition by deploying the concepts refined in the first two sections.

4 Numerical cognition

In this section I outline the phylogenetic basis of mathematical cognition. That basis is in our shared sense of quantity and our ability to estimate the size of small sets by making approximate judgements of the size of the set. This ancient endowment is the basis for our mathematical competence, but it is not all there is to mathematical cognition. This is because precise mathematics depends upon a very recent and acquired public system of exact and discrete mathematical thinking. The ancient system is analogue and approximate, but mathematics requires digital and discrete representations and exact operations. These are, of course, recent additions to inherited cognitive capital. I shall show why mathematical cognition requires our ancient capacity for numerosity and how it is constituted by cognitive practices—which transform our cognitive abilities, resulting in novel and unique modern human cognitive capacities. However, this transformation results in two partially overlapping systems—the approximate number system and the discrete number system —with the latter having unique properties acquired from cultural innovation. One of the puzzles is how it is possible to move from an inherited approximate system to an acquired exact system. The process of enculturation provides the mechanisms by which such a move takes place, from the ancient capacity for numerosity to development in a socio-cultural niche, and the orchestrating role of practices in the assembly of the cognitive systems responsible for mathematical cognition.

26 This section has put together a case for the flexibility of modern minds and the ability to acquire cultural innovations quickly and easily in ontogeny.

4.1 Numerosity in animals and humans

There is strong evidence to suggest that we have a basic analogical and non-linguistic capacity to recognise quantity and number. I think that there is overwhelming evidence for an ancient evolutionary capacity to discriminate cardinality, and to determine in an approximate way the quantity of membership of sets. It is obvious how this capacity, for only very small sets, would be beneficial for activities such as foraging, hunting, and so on.

Recent studies have revealed that the neural populations that code for number are distributed in the intraparietal sulcus (Dehaene & Cohen 2007). A growing number of studies show that both animals and humans possess a rudimentary numerical competence, which is an evolutionary endowment. For example, red-backed salamanders have been shown to choose the larger of two groups of live prey (Uller et al. 2003). Single neuron activation studies in rhesus monkeys (Nieder et al. 2006) discovered that individual neurons respond to changes in number when presented visually (and non-symbolically). These neurons are also located in the intraparietal sulci, indicating a probable cross-species homology. The neurons peak at the presentation of a specific quantity of dots, but then decrease as the numbers presented differ from the original. So a neuron that peaks at the presentation of two dots responds less to three or four dots. The further the numerical distance of the array of dots is from the magnitude to which the neuron is tuned, the lower the firing rate of the neuron. Therefore, the ancient capacity for numerosity is an approximate function, not a discrete one (DeCruz 2008).

This is not yet counting; counting is exact enumeration. Subitizing is the ability to immediately recognise the size, or number, of a small set—usually <4. Most animals subitize, rather than count. Infant humans also appear to be able to subitize (Rouselle & Noël 2008). This ancient or approximate number system (ANS) is a non-linguistic continuous representation[27] of quantities above 4; Dehaene calls it the number sense (1997). Take the following example. Whilst it is easy enough to determine which of the following two boxes contains the larger number of dots without having to count them:

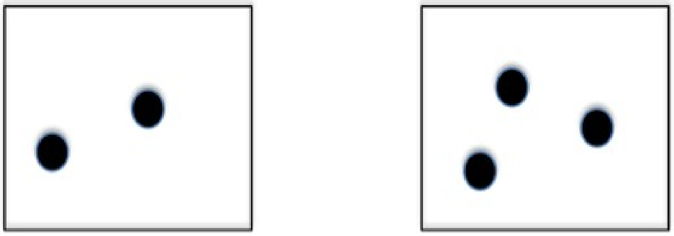

Figure 1: Subitizing or counting?

It is less easy to do so for the following (you will probably need to resort to counting):

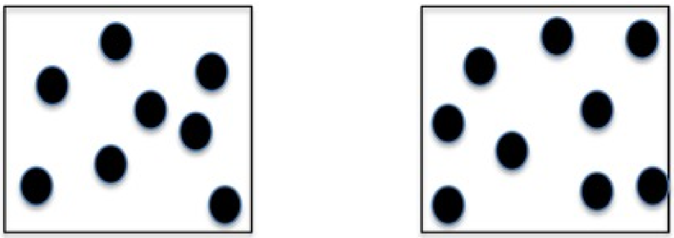

Figure 2: Subitizing or counting?

It is also possible to make estimations or approximate judgements of scale for numbers. Most people can quickly identify that 7 is larger than 3. Even for more complicated exact operations we can do this:

34 + 47 = 268 (is this right?)

We readily reject this result, because the proposed quantity is too distant from the operands of the addition (Dehaene 2001, p. 28).

34 x 47 = 1598 (is this right?)

Approximation involving proximity and distance will not help here (unless you are very practised at mental multiplication), but you might resort to a multiplication algorithm (which might be routinized). It is clear that we have an ancient sense of quantity and are good at making judgements about more than and less than, but when it comes to precise and discrete quantities (particularly larger numbers) we need new capacities to be able to make judgements about operations on discrete numbers.

4.2 Two overlapping systems

The approximate numerical system is an analogue and approximate system for discriminating non-symbolic numerosities greater than 4, but the "representations" are approximate and noisy. The second system is acquired and concerns discrete symbolic and linguistic representation of individual numbers from our numeral system, including individual words for numbers. This system works with discrete, exact, symbolic representations of quantity and allows for the exact operations of arithmetic and mathematics. I will call this the discrete numerical system (DNS). There is disagreement about how much the two systems overlap. However, what is clear is that the internalisation of the public numeral system allows us to perform the kind of digital mathematical operations that are required for most arithmetic and mathematical operations (Nieder & Dehaene 2009, p. 197).

Dehaene and colleagues produced a series of experiments that demonstrate the separate functioning of the two systems. Russian–English bilinguals were taught a set of exact and approximate sums of two digit numbers in one of their languages (Dehaene et al. 1999, p. 970). Their tasks were split into giving exact answers to additions and giving an approximate answer to the addition task. The interesting result was that:

> [w]hen tested on trained exact addition problems, subjects performed faster in the teaching language than in the untrained language, whether they were trained in Russian or English. (Dehaene et al. 1999, p. 971)

This provided evidence that knowledge of arithmetic was being stored in a linguistic format,

27 The appearance of the word representation here need not raise concerns; these are not representations with propositional contents and truth conditions. They are not symbolic and are not molecular constituents that can be combined to make more complex representations.

and that there was a switching cost between the trained and untrained languages. By contrast, there was equivalent performance in the approximation task, and no switching cost between the trained and untrained languages. Dehaene et al. conclude that this provides "evidence that the knowledge acquired by exposure to approximate problems was stored in a language-independent form" (1999, p. 971).

This leads us to the conclusion that there are two overlapping, but not identical, systems for mathematical cognition. The first is the ancient and approximate system, the second is a relatively new and acquired system for discrete and digital representations and operations. As Dehaene & Cohen put it:

> The model that emerges suggests that we all possess an intuition about numbers and a sense of quantities and of their additive nature. Upon this central kernel of understanding are grafted the arbitrary cultural symbols of words and numbers […]. The arithmetic intuition that we inherit through evolution is continuous and approximate. The learning of words and numbers makes it digital and precise. Symbols give us access to sequential algorithms for exact calculations. (2007, p. 41)

The two systems are overlapping but not identical because they have quite different properties. First, the ancient system is part of our phylogeny, whereas the discrete system is an acquired set of capacities in ontogeny. Second, the ancient system is analogue and approximate, whereas the discrete system is digital and exact. Third, the discrete system operates on symbols that don't map directly on to the ancient system.

When we consider very large numbers, such as 10,000,000, there is no obvious analogue in the ANS. Consequently, large or exotic numbers and operations on them do not map onto existing cortical circuitry for numerosity. Lyons et al. (2012) call this phenomenon "symbolic estrangement". Symbols become estranged through a process of symbol-to-symbol mappings, rather than symbol-to-approximate-quantity mappings (Lyons et al. 2012, p. 635).

However, there appears to be a point of contention here: Dehaene expects there to be a more or less direct mapping of symbols to quantities (e.g., the mental number line). If symbolic estrangement does happen, then this would appear to be mistaken. Lyons, Ansari and Beilock propose a developmental resolution of this apparent disagreement. Children may start out in the acquisition of discrete number systems by a mapping to an existing approximate neural coding of quantity, but as the system matures and symbols become abstracted from the ancient system, the mature system splits into two (related but not entirely overlapping) systems: neural circuitry in the DNS tunes for discrete symbols,[28] whereas circuitry in the ANS tunes for approximate quantities, such that discrete symbols do not map directly onto approximate quantities. E.g., 10,000,000. The DNS has properties that are unique.

In the next section I return to the question of the role of practices in assembling the DNS.

5 Mathematical practices

The DNS is dependent upon mathematical practices, systems of number and algorithms for performing mathematical operations, complex mathematical concepts such as sets, functions, and so on. None of these practices, representations, or concepts are innate, and no one seriously thinks that they are. They are culturally inherited and acquired in the right learning niche with experts willing to teach. These new abilities are continuous with our cognitive phylogeny. How, though, can we put the whole package together? This section does that job.

5.1 Cognitive practices and the development of mathematical competence

Mathematics and writing systems are examples of culturally evolved symbol systems that are deployed to complete complex cognitive tasks. These systems are structured by rules and

28 There is evidence of narrower tuning curves for Arabic numerals in the left intraparietal sulcus (Ansari 2008).

norms, but they are deployed as practices: patterns of action spread out across a population. In this case cognitive agents must gain mastery over the symbols, including numerals and operators, as well as the rules for their combination. However, they must also learn how to write and manipulate the symbols according to those rules in order to produce the right products—and this is proceduralised.

There may be more than one way of achieving a solution to the task. One can multiply by the partial products algorithm, or one can use the lattice/grid method or a number of others that have been developed by different cultures using different numerical systems. However, they all involve the same set of features: symbols, rules, operators, spatial configuration, and products, and they jointly constitute a practice for manipulating the symbols to complete mathematical problems. The practices are novel and unique to humans.

The methods apply equally to their off-line equivalents, so in the page-based version of the partial products algorithm we perform the multiplications from right to left and write down their products in rows, carrying numbers where necessary. In the off-line version we can perform the same operations on imagined numerals, multiplying numbers along the line and carrying any numbers as required. It is cognitively taxing to hold the products of the multiplications constant in working memory, though some people can train themselves to become quite good at it. Most people learn off-line multiplication by performing shortcuts; if I want to work out what 25 x 7 is, I just add 25 together 7 times.

On-line methods can change even within the same arithmetical systems, so the partial products algorithm works like this:

```
     23
x    11
     23   (1x3 and 1x2)
+   230   (carry 0, 1x2 and 1x3)
    253   (add products together)
```

However there is an equivalent algorithm that works like this:

```
     23
x    11
    200   (10 x 20)
     30   (10 x 3)
+    23   (1 x23)
    253   (add products together)
```

The algorithms may differ, but they still involve the practice of spatially arranging the numerals, and performing operations on them and deriving a product, by performing the staged manipulations on the page. It appears then to matter how we manipulate symbols in public space, but is there any empirical evidence for this conclusion?

CI predicts that it matters how symbols are spatially arranged when they are being manipulated. Landy & Goldstone (2007) found that college-level algebraists could be induced to make errors by altering the layout of numbers that they were to manipulate. They did this by altering the spacing of the equations:

F+z * t+b = z+f * b+t

Although minor, the extra spacing was enough to induce errors. It matters how the symbols are spatially laid out, for this layout is the basis of how we manipulate those symbols. In this case the artificial visual groups created by the irregular spacing affected the judgement of the validity of the equation. If the visual groupings were inconsistent with valid operator precedence then they negatively affected the judgement.[29]

Landy & Goldstone's work provides evidence that expert algebraists are practised at symbolic reasoning achieved via the perception and manipulation of physical notations (2007; Landy et al. 2014). Rather than an internal system of abstract symbols and rules for their combination (i.e., a language of thought), the system is composed of perceptual-motor systems and the manipulations of numerals. They are careful to say that the manipulations must conform to the abstract norms of algebra. Dutilh Novaes (2013) takes this to be evidence that mathematical competence is constituted by the

29 In algebra multiplications are made before additions. E.g., 5+2*6 = 17 (not 42).

capacity to manipulate inscriptions of mathematical equations. This fits very well with the CI approach.

Despite some interesting lacunae (savants and blind mathematicians), most mathematicians learn to manipulate numerals and other mathematical symbols on the page, and they continue to do so throughout their mature cognitive lives. Landy and Goldstone's evidence supports the thesis that mathematical competence is constituted, in part, by our capacity to manipulate symbols in public space; that competence is, properly, a matter of interaction.

5.2 Continuity and transformation

We have seen that there is an ancient evolutionary endowment for numerosity—an analogue and approximate system. This system is found in other primates and other species. It provides both the phylogenetic basis of mathematical cognition and the initial constraints for the development of the DNS. The DNS did not spring sui generis into the world. It did so because of a heady mixture of socio-cultural pressures, phenotypic and neural plasticity, social learning strategies, and cultural inheritance. These are the conditions for the scaffolding of the ANS, transforming our basic biological capacities into the DNS.

New cultural functions, discrete mathematical functions, and the practices for manipulating inscriptions transform existing circuitry in the brain. Once we learn how to recognise, understand, and manipulate mathematical symbols our brains undergo a profound transformation. There is a reproducible circuit for mathematical cognition involving a bi-lateral parietal based approximate estimation; a left lateralised verbal framework for arithmetic concepts (e.g., number words); and a occipito-temporal based symbol recognition system (e.g., Arabic numerals). The system also incorporates visual-motor systems for writing (manipulating, or pushing) symbols in public space.

A further important aspect of transformation is symbolic estrangement. As the DNS matures it becomes more abstract and less directly mapped onto the approximate functions of the ANS. Interestingly, at the same time expert mathematicians become reliant upon visual-motor capacities for manipulating inscriptions. Transformation depends upon the novelty and uniqueness of mathematical symbols and practices.

5.3 Novelty and uniqueness

Symbolic number systems and sequential algorithms allow for mathematical and cognitive novelty. Once we have a public system, all manner of exotic numbers and operations can be discovered:[30] negative numbers, square roots, zero, sets, and so on. Its importance lies in the ability to perform computations that cannot be performed by ancient neural functions for numerosity. For example, the neural circuits responsible for numerosity cannot (on their own) represent -3 or $\sqrt{54}$, and yet this is simply represented in terms of public mathematical symbols (DeCruz 2008). This is because the symbolic representations are novel and unique. Initially, novelty results from the pressures of increasing social and economic complexity. Small roaming bands of foragers do not need to develop symbolic number systems; post-agricultural Neolithic societies settled in villages and towns do. A further issue is how novelty comes about from the ability to abstractly combine symbols and functions that apply to the symbols. I don't propose to try to answer that question here; however, we might think of this as a curiosity- and creativity-driven processes. Given uniquely human behavioural and neural plasticity and socio-cultural complexity we might expect an increasing drive towards cognitive innovation. This has certainly been the story of recent cultural evolution in modern human societies.

This concludes the discussion of mathematical cognition as enculturation. Now I turn to the objections.

6 The incredible shrinking system

Why not just shrink the cognitive system to brain-based systems? Is there a way to bridge

30 I will not address the issue of what discovery amounts to here and will remain neutral on whether discovery reveals a platonic mathematical system or simply the logical relations between concepts.

the impasse between moderate and strong embedding? One argument concerns whether it makes any difference to cognitive science to consider, for example, the manipulation of public symbols to be cognitive processes (Sprevak 2010). Ultimately, to give a decisive answer to that question we would need to change our conception of cognitive processes to on-going dynamical interactions with the environment that loop through brain, body, and environment. However, weak and moderate embedded approaches do not work with such a conception of cognitive process; they work with an input-process-output style sandwich model, where processes supervene on bodily states and processes. For them, there is no reason to accept strong embedding, and much of the discussion has been based around thought experiments or abstract definitions rather than concrete examples.

However, even on a scaffolded view of cognition we can't deny the difference-making role the manipulations of symbols make to the completion of cognitive tasks. Manipulating public symbols is unique; there is a difference between internalised strategies for completing mathematical tasks and strategies for manipulating mathematical inscriptions. Our cognitive capacities cannot cope with long sequences of complex symbols and operations on them. This is why we must learn strategies and methods for writing out proofs. Symbol manipulation makes a unique difference to our ability to complete mathematical tasks, and we cannot simply ignore their role. If we take the approach of CI, then mathematical cognition is constituted by these bouts of symbol manipulation, and we cannot simply shrink the system back to the brain. The case for a strongly embedded approach to mathematical cognition depends upon the novelty and uniqueness of mathematical practices and dual component transformations. Our evolutionary endowments of numerosity are not up to the task of exact symbolic arithmetic and mathematics. Without symbolic number systems and sequential algorithms there would be no mathematical innovation. Mathematical innovation includes representational novelty: negative numbers, square roots, zero, etc., but also novel functions: multiplication, division, etc. Novelty comes about from the ability to abstractly combine symbols and functions that apply to the symbols.

Uniquely, symbols represent quantities discretely, but there is also the unique human capacity of manipulating symbols in public space. We learn to manipulate symbols in public space and we continue to do so when completing cognitive tasks.

The entire system of mathematics is not contained in a single brain. Symbol systems are public systems of representations and practices for their manipulation. Mathematical practices are part of the niche that we inherit—they are part of our cultural inheritance.

6.1 Impermanent scaffolds?

Another objection concerns the impermenance of the scaffolding required for mathematical cognition. Once we have internalised the scaffolding of symbolic number systems, we have no further need for it, except for communication purposes. This claim would be proven if we did not continue to manipulate numerals when completing cognitive tasks. Even if we think that transformation only results in new internal representational resources, and that this just amounts to moderate embedding/scaffolding, we must also concede that most mathematics is conducted on the page.

Scaffolding theorists, like Sterelny, can endorse this idea; indeed they can agree with the bulk of the framework provided by CI whilst avoiding the constitutive claim. What they cannot do is deny that mathematical practice and the manipulation of physically laid-out symbols on the page is a difference maker for mathematical cognition. If you remove it, the ability to complete mathematical tasks drops considerably. To do so is to fly in the face of the empirical evidence from psychology (Landy & Goldstone 2007) and cognitive neuroscience (Dehaene & Cohen 2007; Ansari 2012). Consequently, it is clear that cognitive practices transform our mathematical abilities, lending weight to the CI approach.

The case I have presented in this paper is that symbols are not simply impermanent scaffolds, they are permanent scaffolds. They become part of the architecture of cognition (and not simply through internalisation). Mastery of symbol systems results in changes to cortical circuitry, altering function and sensitivity to a new, public, representational system. However, it also results in new sensori-motor capacities for manipulating symbols in public space. The case can be made in terms of what a symbol system is:

A symbol is a physical mark (or trace), either in physical space, or as a digital trace. Symbol systems contain rules and practices for interpreting symbols, for combining them, and for ordering and manipulating them. A large body of often tacit practices for interpreting and manipulating symbols is acquired. Scaffolding is not simply an amodal symbol with an abstract designation that needs to be learnt (or mapped onto some innate symbol); scaffolding is also how the symbols are physically arranged, how symbols are pushed from one place to the next in a regular fashion. Finally, scaffolding is also how we use our own bodies, eyes, ears, and hands to create and manipulate symbols.

7 Conclusion

I have presented a case for CI as a process of enculturation, with mathematical cognition as an example of the process of enculturation at work. I began by laying out the 4E landscape and locating CI within it, relative to enactivism and EM. In particular I showed how CI shares the interactive stance of enactivism and the constitutive stance of EM, but how it also differs from these. The main difference between CI and enactivism is that CI does not equate life and mind in the way that enactivism does. The main difference between CI and EM is that CI takes cultural practices to play a central role in the assembly of cognitive systems, whereas EM does not.

I then went on to outline the central concepts required to make sense of enculturation. The CI framework embraces both evolutionary continuity and transformation of existing cognitive circuitry in development. Our modern minds are built on archaic precursors by slow incremental changes. However, modern humans are behaviourally plastic and scaffolded learning drives functional changes in our plastic brains. The developmental change from the ANS to the DNS is an example of how learning-driven changes to cortical function result in new abilities, but this would not happen without the novelty and uniqueness of mathematical symbols and the practices for manipulating them.

I also countered two standard objections: impermanence and shrinkage. The defence of CI rested on the novelty and uniqueness of mathematical practices and symbols.

If the CI framework is on the right track, then human cognitive evolution has resulted in minds that are flexible and interactive. Furthermore, cultural evolution has resulted in written symbol systems and practices for manipulating symbols that can be acquired (in development) by minds like ours. The uniqueness of modern human minds lies in their capacity for transformation.

References

Adams, A. & Aizawa, K. (2008). *Defending the bounds of cognition.* Oxford, UK: Blackwell.

Andrews, K. (2012). *Do apes read minds?: Toward a new folk psychology.* Cambridge, MA: MIT Press.

Ansari, D. (2008). Effects of development and enculturation on number representation in the brain. *Nature Reviews Neuroscience*, *9* (4), 278-291. 10.1038/nrn2334

——— (2012). Culture and education: New frontiers in brain plasticity. *Trends in Cognitive Sciences*, *16* (2). 10.1016/j.tics.2011.11.016

Barbaras, R. (2010). Life and exteriority: The problem of metabolism. In J. Stewart, O. Gapenne & E. Di Paolo (Eds.) *Enaction toward a new paradigm for cognitive science* (pp. 89-122). Cambridge, MA: MIT Press.

Barkow, J. H., Cosmides, L. & Tooby, J. (1992). *The adapted mind: Evolutionary psychology and the generation of culture.* Oxford, UK: Oxford University Press.

Beer, R. D. (2000). Dynamical approaches to cognitive science. *Trends in Cognitive Sciences*, *4* (3), 91-99. 10.1016/S1364-6613(99)01440-0

Boyd, R. & Richerson, P. J. (2005). *The origin and evolution of cultures.* Oxford, UK: Oxford University Press.

Clark, A. (2008). *Supersizing the mind: Embodiment, action, and cognitive extension.* Oxford, UK: Oxford University Press.

——— (2011). Finding the mind. *Philosophical Studies*, *152* (3), 447-461. 10.1007/s11098-010-9598-9

——— (2015). Embodied prediction. In T. Metzinger & J. M. Windt (Eds.) *Open MIND.* Frankfurt a. M., GER: MIND Group.

Davies, S. (2012). *The artful species: Aesthetics, art, and evolution.* Oxford, UK: Oxford University Press.

De Cruz, H. (2008). An extended mind perspective on natural numberrepresentation. *Philosophical Psychology*, *21* (4), 475-490. 10.1080/09515080802285289

Dehaene, S. (1997). *The number sense: How the mind creates mathematics.* London, UK: Penguin.

——— (2001). Précis of the number sense. *Mind & Language*, *16* (1), 16-36.

——— (2009). *Reading in the brain: The new science of how we read.* London, UK: Penguin.

Dehaene, S. & Cohen, L. (2007). Cultural recycling of cortical maps. *Neuron*, *56* (2), 384-398. 10.1016/j.neuron.2007.10.004

——— (2011). The unique role of the visual word form area in reading. *Trends in Cognitive Sciences*, *15* (6), 254-262. 10.1016/j.tics.2011.04.003

Dehaene, S., Spelke, E., Pinel, P., Stanescu, R. & Tsivkin, S. (1999). Sources of mathematical thinking: Behavioral and brain-imaging evidence. *Science*, *284* (5416), 970-974. 10.1126/science.284.5416.970

Dutilh Novaes, C. (2013). Mathematical reasoning and external symbolic systems. *Logique & Analyse*, *56* (221), 45-65.

Finlayson, C. (2009). *The humans who went extinct: Why Neanderthals died out and we survived.* Oxford, UK: Oxford University Press.

Gallagher, S. (2005). *How the body shapes the mind.* Oxford, UK: Oxford University Press.

Gallese, V. (2008). Mirror neurons and the social nature of language: The neural exploitation hypothesis. *Social Neuroscience*, *3* (3-4), 317-333. 10.1080/17470910701563608

Glenberg, A. (2010). Embodiment as a unifying perspective for psychology. *Cognitive Science*, *1* (4), 586-596. 10.1002/wcs.55

Godfrey-Smith, P. (1996). *Complexity and the function of mind in nature.* Cambridge, UK: Cambridge University Press.

Hoffman, M. (2014). Evolution of the human brain: When bigger is better. *Frontiers in Neuroanatomy*, *8* (1). 10.3389/fnana.2014.00015

Huebner, B. (2013). *Macrocognition: Distributed minds and collective intentionality.* New York, NY: Oxford University Press.

Hurley, S. (2010). The varieties of externalism. In R. Menary (Ed.) *The extended mind* (pp. 101-154). Cambridge, MA: MIT Press.

Hutchins, E. (1995). *Cognition in the wild.* Cambridge, MA: MIT Press.

——— (2011). Enculturating the supersized mind. *Philosophical Studies*, *152* (3), 437-446. 10.1007/s11098-010-9599-8

Huttenlocher, P. R. (2002). *Neural plasticity: The effects of environment on the development of the cerebral cortex.* Cambridge, MA: Harvard University Press.

Hutto, D. (2008). *Folk psychological narratives: The sociocultural basis of understanding reasons.* Cambridge, MA: MIT Press.

Hutto, D. D. & Myin, E. (2013). *Radicalizing enactivism: Basic minds without content.* Cambridge, MA: MIT Press.

Jeffares, B. (2010). The co-evolution of tools and minds: Cognition and material culture in the hominin lineage. *Phenomenology and the Cognitive Sciences*, *9* (4), 503-520. 10.1007/s11097-010-9176-9

Laland, K. N., Odling-Smee, J. & Feldman, M. W. (2000). Niche construction, biological evolution and cultural change. *Behavioral and Brain Sciences*, *23* (1), 131-146. 10.1017/S0140525X00002417

Landy, D., Allen, C. & Zednik, C. (2014). A perceptual account of symbolic reasoning. *Frontiers in Psychology*, *5* (275). 10.3389/fpsyg.2014.00275

Landy, D. & Goldstone, R. L. (2007). How abstract is symbolic thought? *Journal of Experimental Psychology*, *33* (4), 720-733. 0.1037/0278-7393.33.4.720

Lycett, S. J. & Gowlett, J. A. J. (2008). On questions surrounding the Acheulean "tradition". *World Archaeology*, *40* (3), 295-315. 10.1080/00438240802260970

Lyons, I. M., Ansari, D. & Beilock, S. L. (2012). Symbolic estrangement: Evidence against a strong association between numerical symbols and the quantities they represent. *Journal of Experimental Psychology: General*, *141* (4), 635-641. 10.1037/a0027248

MacLean, E. L., Matthews, L. J., Hare, B. A., Nunn, C. L., Anderson, R. C., Aureli, F., Brannon, E. M., Call, J., Drea, C. M., Emery, N. J., Haun, D. B., Herrmann, E., Jacobs, L. F., Platt, M. L., Rosati, A. G., Sandel, A. A., Schroepfer, K. K., Seed, A. M., Tan, J., van Schaik, C. P. & Wobber, V. (2012). How does cognition evolve? Phylogenetic comparative psychology. *Animal cognition*, *15* (2), 223-238. 10.1007/s10071-011-0448-8

Menary, R. (2007). *Cognitive integration: Mind and cognition unbounded.* London, UK: Palgrave Macmillan.

——— (2010). Dimensions of mind. *Phenomenology and the Cognitive Sciences*, *9*, 561-578. 10.1007/s11097-010-9186-7

——— (2011). Our glassy essence: The fallible self in pragmatist thought. In S. Gallagher (Ed.) *The Oxford Handbook of the Self* (pp. 609-632). Oxford, UK: Oxford University Press.

——— (2012). Cognitive practices and cognitive character. *Philosophical Explorations: An International Journal for the Philosophy of Mind and Action*, *15* (2), 147-164. 10.1080/13869795.2012.677851

——— (2014). Neuronal recycling,neural plasticity and niche construction. *Mind and Language*, *29* (3), 286-303. 10.1111/mila.12051

Menary, R. & Kirchhoff, M. (2014). Cognitive transformations and extended expertise. *Educational Philosophy and Theory*, *46* (6), 610-623. 10.1080/00131857.2013.779209

Nieder, A. & Dehaene, S. (2009). Representation of number in the brain. *Annual Review of Neuroscience*, *32*, 185-208. 10.1146/annurev.neuro.051508.135550

Nieder, A., Diester, I. & Tudusciuc, O. (2006). Temporal and spatial enumeration processes in the primate parietal cortex. *Science*, *313* (5792), 1432-1435. 10.1126/science.1130308

Odling-Smee, F. J., Laland, K. N. & Feldman, M. F. (2003). Niche construction: The neglected process in evolution. *Monographs in Population Biology*, *37*

Price, C. J. & Devlin, J. T. (2011). The interactive account of ventral occipitotemporal contributions to reading. *Trends in Cognitive Sciences*, *15* (6), 246-253. 10.1016/j.tics.2011.04.001

Roepstorff, A., Niewöhner, J. & Beck, S. (2010). Enculturing brains through patterned practices. *Neural Networks*, *23* (8-9), 1051-1059.

Rouselle, L. & Noël, M. P. (2008). The development of automatic numerosity processes in preschoolers: Evidence for numerosity-perceptual interference. *Developmental Psychology*, *44* (2), 544-560. 10.1037/0012-1649.44.2.544

Rowlands, M. (2010). *The new science of the mind: From extended mind to embodied phenomenology.* Cambridge, MA: MIT Press.

Rupert, R. (2009). *Cognitive systems and the extended mind.* Oxford, UK: Oxford University Press.

Semaw, S., Rogers, M. J., Quade, J., Renne, P. R., Butler, R. F., Dominguez-Rodrigo, M., Stout, D., Hart, W. S., Pickering, D. & Simpsons, S. W. (2003). 2.6-Million-year-old stone tools and associated bones from OGS-6 and OGS-7, Gona, Afar, Ethiopia. *Journal of Human Evolution*, *45* (2), 169-177. 10.1016/S0047-2484(03)00093-9

Shultz, S., Nelson, E. & Dunbar, R. I. M. (2012). Hominin cognitive evolution: Identifying patterns and processes in the fossil and archaeological record. *Philosophical Transactions of the Royal Society*, *367* (1599), 2130-2140. 10.1098/rstb.2012.0115

Smart, J. J. C. (1959). Sensations and brain processes. *Philosophical Review*, *68*, 141-156.

Sprevak, M. (2010). Inference to the hypothesis of extended cognition. *Studies in History and Philosophy of Science Part A*, *41* (4), 353-362. 10.1016/j.shpsa.2010.10.010

Sterelny, K. (2003). *Thought in a hostile world : The evolution of human cognition.* Oxford, UK: Blackwell.

——— (2010). Minds: Extended or scaffolded? *Phenomenology and the Cognitive Sciences*, *9* (4), 465-481. 10.1007/s11097-010-9174-y

——— (2012). *The evolved apprentice : How evolution made humans unique.* Cambridge, MA: MIT Press.

Stewart, J. R., Gapenne, O. & Di Paolo, E. A. (2010). *Enaction: Toward a new paradigm for cognitive science.* Cambridge, MA: MIT Press.

Stich, S. (1983). *From folk psychology to cognitive science: The case against belief.* Cambridge, MA: MIT Press.

Stout, D., Toth, N., Schick, K. & Chaminade, T. (2008). Neural correlates of Early Stone Age toolmaking: Technology, language and cognition in human evolution. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *363* (1499), 1939-1949. 10.1098/rstb.2008.0001

Thatcher, R. W. (1991). Maturation of the human frontal lobes: Physiological evidence for staging. *Developmental Neuropsychology*, *7* (3), 397-419. 10.1080/87565649109540500

Theiner, G. (2013). Onwards and upwards with the extended mind: From individual to collective epistemic action. In L. Caporael, J. Griesemer & W. Wimsatt (Eds.) *Developing scaffolds* (pp. 191-208). Cambridge, MA: MIT Press.

Thompson, E. (2007). *Mind in life: Biology, phenomenology, and the sciences of mind.* Cambridge, MA: Harvard University Press.

Toth, N. & Schick, K. D. (2006). *The Oldowan: Case studies into the earliest stone age.* Bloomington, IN: Stone Age Institute Press.

Turner, J. S. (2000). *The extended organism: The physiology of animal-built structures.* Cambridge, MA: Harvard University Press.

Uller, C., Jaeger, R., Guidry, G. & Martin, C. (2003). Salamanders (Plethodon cinereus) go for more: Rudiments of number in an amphibian. *Animal Cognition*, *6* (2), 105-112. 10.1007/s10071-003-0167-x

Varela, F., Thompson, E. & Rosch, E. (1991). *The embodied mind.* Cambridge, MA: MIT Press.

Vygotsky, L. (1981). The instrumental method in psychology. In J. Wertsch (Ed.) *The concept of activity in soviet psychology.* Armonk, NY: Sharpe.

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

Wheeler, M. & Clark, A. (2008). Culture, embodiment and genes: Unravelling the triplehelix. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *363* (1509), 3563-3575. 10.1098/rstb.2008.0135

Whiten, A., Hinde, R. A., Laland, K. N. & Stringer, C. B. (2011). Culture evolves. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *366* (1567), 938-948. 10.1098/rstb.2010.0372

Enriching the Notion of Enculturation: Cognitive Integration, Predictive Processing, and the Case of Reading Acquisition

A Commentary on Richard Menary

Regina E. Fabry

Many human cognitive capacities are rendered possible by enculturation in combination with specific neuronal and bodily dispositions. Acknowledgment of this is of vital importance for a better understanding of the conditions under which sophisticated cognitive processing routines could have emerged on both phylogenetic and ontogenetic timescales. Subscribing to enculturation as a guiding principle for the development of genuinely human cognitive capacities means providing a description of the socio-culturally developed surrounding conditions and the profound neuronal and bodily changes occurring as a result of an individual's ongoing interaction with its cognitive niche. In this commentary, I suggest that the predictive processing framework can refine and enrich important assumptions made by the theory of cognitive integration and the associated approach to enculturated cognition. I will justify this suggestion by considering several aspects that support the complementarity of these two frameworks on conceptual grounds. The result will be a new integrative framework which I call enculturated predictive processing. Further, I will supplement Richard Menary's enculturated approach to mathematical cognition with an account of reading acquisition from this new perspective. In sum, I argue in this paper that the cognitive integrationist approach to enculturated cognition needs to be combined with a predictive processing style description in order to provide a full account of the neuronal, bodily, and environmental components giving rise to cognitive practices. In addition, I submit that the enculturated predictive processing approach arrives at a conceptually coherent and empirically plausible description of reading acquisition.

Keywords
Cognitive integration | Cognitive transformation | Enculturation | Neural plasticity | Neuronal reuse | Predictive processing | Reading acquisition | Scaffolded learning

Commentator

Regina E. Fabry
fabry@students.uni-mainz.de
Johannes Gutenberg-Universität
Mainz, Germany

Target Author

Richard Menary
richard.menary@mq.edu.au
Macquarie University
Sydney, Australia

Editors

Thomas Metzinger
metzinger@uni-mainz.de
Johannes Gutenberg-Universität
Mainz, Germany

Jennifer M. Windt
jennifer.windt@monash.edu
Monash University
Melbourne, Australia

1 Introduction

In his target paper *Mathematical Cognition: A Case of Enculturation*, Richard Menary investigates the conditions under which phylogenetically recent, socio-culturally shaped target phenomena within cognitive science such as mathematics, reading, and writing have emerged. Resting on his theory of cognitive integration (CI; e.g., Menary 2007a), he starts from the idea that these processes are fully continuous with phylogenetically older ones (*evolutionary continuity*). This type of continuity is justified by the assumption that the evolution of neuronal reuse mechanisms allows for the redeployment of cortical circuits for phylogenetically recent functions (Anderson 2010; Anderson & Finlay 2014). Ontogenetically, neuronal reuse is a precondition of *learning driven plasticity* (LDP), which “can result in both structural and functional changes in the brain” (Menary this collection, p. 8). That is, the human brain is assumed to be neuronally plastic so that its processing routines are altered as the individual acquires new cognitive abilities (Ansari 2012). However, the acquisition of new cognitive abilities takes place within

> […] a highly structured cognitive niche that contains not only physical artefacts, but also: representational systems that embody knowledge (writing systems, number systems, etc.); skills and methods for training and teaching new skills (Menary & Kirchhoff 2014); practices for manipulating tools and representations. (Menary this collection, p. 6)

It is this cognitive niche that provides the resources for *scaffolded learning*, which allows the individual to acquire new cognitive abilities through its ongoing embodied interaction with its socio-cultural environment. Together, LDP and scaffolded learning lead to cognitive transformations that augment the individual’s cognitive capacities through ontogenesis: “Cognitive transformations result from our evolved plasticity and scaffolded learning in the developmental niche” (Menary this collection, p. 8).[1] The result of cognitive transformation is the acquisition of a sufficient degree of expertise in performing a certain *cognitive practice*. Cognitive practices are normatively constrained to the extent that socio-culturally shaped procedures work in close interaction with the cognitive niche: They “[…] are culturally endowed (bodily) manipulations of informational structures” (Menary this collection, p. 4), such as manipulations of tokens of a representational writing system, and they serve to complete a cognitive task. In order to describe the transformational processes by which cognitive practices are acquired, Menary introduces the notion of *enculturation*: “Enculturation rests on the acquisition of cultural practices that are cognitive in nature” (ibid.). That is, enculturation refers to any cognitive transformation that is rendered possible by LDP and the individual’s ongoing interaction with its cognitive niche. As a proof of concept, Menary (this collection) deals with mathematical cognition and describes the ways in which individuals acquire expertise in manipulating a public, socio-culturally developed mathematical symbol system. Relying on a set of empirical results, he arrives at the conclusion that precise mathematical operations are rendered possible by the recruitment of a neuronal sub-system during ontogeny. In contrast to the evolved approximate number system (ANS), which allows for subitizing and is also present in other animals, the neuronal realization of the discrete number system (DNS) heavily depends on LDP, the individual’s immersion into its cognitive niche, and its active participation in scaffolded learning routines. Thus, the acquisition of mathematical skills is an important example of enculturation.

The purpose of this commentary is to enrich and refine the enculturated approach. First, I will propose that the predictive processing framework provides conceptual and explanatory tools for describing and explaining the neuronal and extracranial bodily mechanisms underlying cognitive practices and enculturation. Thus, I will accept the challenge to combine “[…] the dynamical nature of causal commerce between world, body, and brain and the inferential free energy principle that allows their unification in one account” (Hohwy this collection, p. 18). I will argue that a new integrative framework that views CI and predictive processing as complementary is able to meet this challenge. Second, I will illustrate this by presenting reading acquisition as a paradigmatic case of enculturated cognition. In particular, I will demonstrate that a position that combines the enculturated approach with predictive processing,

1 More precisely, according to Menary (2014, p. 293) it is scaffolded learning that renders LDP possible in the course of cognitive development of individuals: “Both structural and functional plasticity can result from both endogenous and exogenous sources, but here the focus is on structural and functional changes driven by scaffolded learning.”

which I call enculturated predictive processing, leads to a parsimonious and conceptually coherent account of reading acquisition that helps interpret and unify a vast array of recent empirical findings.

2 Towards a more complete approach to enculturation: Cognitive integration and predictive processing

In order to appreciate the descriptive power of the enculturated approach, it is necessary to specify the mechanistic underpinnings of the acquisition of cognitive practices. In his summary of the CI framework, Menary (this collection, p. 2) argues that "[a]lthough the framework is unified by a dynamical systems description of the evolution of processing in the hybrid and multi-layered system, it recognises the novel contributions of the distinct processing profiles of the brain, body and environment." However, the dynamical systems style approach to the acquisition and enactment of cognitive practices in the version first introduced in Menary (2007a, pp. 42-48) does not exhaustively specify the distinct, yet highly interactive neuronal and bodily components of cognitive processing. Furthermore, it does not account for LDP, simply because it remains neutral to the concrete realization of its neuronal component system. Finally, the dynamical systems approach, on Menary's construal, helps illustrate what the interactive contribution of neuronal and extracranial bodily components to human cognition might amount to. Yet, it does not spell out the mutual influence that neuronal and extracranial bodily components have over each other.

This is where predictive processing (PP) enters the picture. In the remainder of this commentary I will argue that the PP approach provides the resources for a more detailed account of how human cognitive systems become enculturated and how they are subject to integrated cognition.

2.1 Cognitive integration: Five theses about human cognition

In its original version (cf. Menary 2007a), CI is constituted by five theses. They emphasize the different aspects that are crucial for an integrationist approach to cognitive processing: 1. Human cognition is continuous with animal cognition on both diachronic and synchronic scales. However, it has a special status in that it is situated in a particular cognitive niche and heavily rests upon neural plasticity which is itself an adaptation (*continuity thesis*). 2. Certain cognitive processes are hybrid because they are constituted by neuronal and extracranial bodily components (*hybrid mind thesis*). 3. In the course of ontogenetic hybrid cognitive processing, both the constitutive neuronal and extracranial bodily functions are transformed (*transformation thesis*). 4. The bodily manipulation of specific environmental resources plays a crucial functional role in integrated cognitive processes (*manipulation thesis*). 5. These manipulations are constrained by cognitive norms, which are acquired through learning, and which realize socio-culturally developed habits for the interaction with cognitive resources (*cognitive norms thesis*).

In addition to the continuity thesis and the cognitive transformation thesis, which are given centre stage in Menary's target paper, the hybrid mind thesis is important in that it acknowledges the close interaction of neuronal and extra-neuronal bodily sub-processes in the completion of cognitive tasks. In other words, certain cognitive processes "involve the integration of neural manipulations of vehicles and bodily manipulations of environmental vehicles" (Menary 2010, p. 236; see also Menary 2007b, p. 627). The notion of bodily manipulation as it is used here goes back to Mark Rowlands' (1999, pp. 23f) account of *environmentalism*, which claims that "cognitive processes are, in part, made up of manipulation of relevant structures in the cognizer's environment". In this context, manipulation is defined as "any form of bodily interaction with the environment – manual or not, intrusive or otherwise – which makes use of the environment in order to accomplish a given task" (ibid., p. 23). Thus, subscribing to the manipulation thesis amounts to the assumption that "[c]ognitive processing often involves these online bodily manipulations of the cognitive niche, sometimes as individuals and sometimes in collaboration with others" (Menary this collection, p. 3). Importantly, it is assumed that extracranial bodily manipulations causally interact with neural sub-processes, thereby stressing the hybridity of cognitive processes (cf. Menary 2007a, p. 138). In addition to highlighting the constitutive role of embodied engagements with "external" cognitive resources as proposed by Rowlands (1999), cognitive integrationists claim that the manipulation of these resources is constrained by cognitive norms. In this vein, Menary (2007a, p. 5; 2010, p. 233) argues that "[o]ur abilities to manipulate the extrabodily environment are normative and are largely dependent on our learning and training histories." The idea that certain cognitive abilities are normatively structured thus concerns the individual's interaction with specific resources provided by the cognitive niche. Importantly, the normatively constrained ways in which environmental resources are integrated into cognitive processes are shared by many individuals. Put differently, the normativity of cognitive practices helps "[…] stabilise and govern interactive thought across a population of similar phenotypes" (Menary this collection, p. 4). Furthermore, the acquisition of a certain cognitive practice is tightly connected with the acquisition of the relevant cognitive norms in the course of scaffolded learning. This is because "we learn cognitive practices by learning the cognitive norms that govern the manipulation of vehicles" (Menary 2007b, p. 628).

From these five theses defended by CI it follows that there should be two distinct, yet interdependent levels of description for cognitive practices. First, there is the social level of description. On this level, cognitive practices need to be approached by highlighting the interactive, cooperative cognitive achievements of a large group of individuals sharing the same cognitive niche. Second, cognitive practices can be investigated by approaching them on an individual level of description. In this case, the acquisition and enactment of a certain cognitive practice is described with regards to a certain individual. However, any individual level description needs to acknowledge that certain cognitive capacities of an enculturated individual are rendered possible only by the individual's ongoing interaction with its socio-culturally shaped environment in normatively constrained ways. This means to do justice to the broader socio-cultural context of enculturated cognition, while being interested in a precise description of its neuronal and extracranial bodily sub-components. In this commentary I will operate on the individual level of description without denying that it is important to develop a fine-grained description on the social level by specifying the properties of a certain cognitive niche and the conditions under which it could have emerged.

To this end, I will now proceed by summarizing the most important features of the predictive processing (PP) approach that will help specify the mechanistic underpinnings of enculturated cognition.

2.2 An outline of predictive processing

Recently, the idea that human perception, action, and cognition can be described and explained in terms of hierarchically organized predictive processing mechanisms implemented in the human brain has enjoyed widespread attention within cognitive neuroscience (e.g., Friston 2005, 2010; Friston et al. 2012), philosophy of mind, and philosophy of cognitive science (e.g., Clark 2012, 2013, this collection; Hohwy 2011, 2012, 2013, 2014, this collection; Seth this collection). The overall epistemic goal of this emerging approach is to describe perceptual, sensorimotor, and cognitive target phenomena within a single framework by relying on unifying mechanistic principles. Accounts of PP generally assume that human perception, action, and cognition are realized by Bayesian probabilistic generative models implemented in the human brain. Since the human brain does not have immediate access to the environmental causes of sensory effects, it has to infer the most probable state of affairs in the environment giving rise to sensory data (cf. Seth this collection, pp. 4f). PP approaches solve this *inverse problem* by assuming that generative models in accordance with Bayes' rule are implemented in the human brain. On this construal, a generative model

"[…] aims to capture the statistical structure of some set of observed inputs by tracking […] the causal matrix responsible for that very structure" (Clark 2013, p. 182). In order to be able to infer the causes of sensory effects, generative models encode probability distributions. Each generative model provides several hypotheses about the causes of a certain sensory input. The system has somehow to 'decide' which hypothesis needs to be chosen in order to account for the cause of the sensory effect. The descriptive power of Bayes' rule lies in its capacity to capture the probabilistic estimations underlying these choices. Applied to the case of human perception, action, and cognition, Bayesian generative models are assumed to be realized in hierarchically organized structures comprising multiple, highly interactive low- and high-level cortical areas. This is referred to as the *Bayesian brain hypothesis* (cf. Friston 2010, p. 129). The hierarchical organization of probabilistic generative models is combined with a specific version of *predictive coding*, where predictive coding "depicts the top-down flow as attempting to predict and fully 'explain away' the driving sensory signal, leaving only any residual 'prediction errors' to propagate forward within the system" (Clark 2013, p. 182). That is to say, selected hypotheses inform prior predictions about the sensory input to be expected at each level of the hierarchy. These predictions fulfil the function of encoding knowledge about statistical regularities of patterns in the observable (or any imaginable) world. This hypothesis selection proceeds in accordance with Bayes' rule. The processing of sensory input gives rise to prediction errors. Prediction errors carry neuronally realized information about "[…] residual differences, at every level and stage of processing, between the actual current signal and the predicted one" (Clark this collection, p. 4). Importantly, it is only prediction errors, and not sensory input *per se,* that are fed forward within the hierarchy (cf. Clark 2013, pp. 182f; Hohwy 2012, p. 3, 2013, p. 47, 2014, p. 4). The overall aim of this multi-level processing mechanism is to *minimize prediction error*, that is, to reduce or to 'explain away' the discrepancy between predictions and the actually given sensory input that is an effect of environmental (or bodily) causes (cf. Clark 2013, p. 187; Hohwy 2011, p. 269, 2013, p. 88). This is known as *prediction error minimization.*[2]

Prediction error minimization is a special way of minimizing *free energy* in accordance with the principle "that any self-organizing system that is at equilibrium with its environment must minimize its free energy" (Friston 2010, p. 127). Applied to human perception, cognition, and action, minimizing free energy means minimizing the amount of unbound energy available to the perceiving, cognizing, and acting organism. This is where prediction error enters the picture. As Andy Clark (2013, p. 186) puts it, "[p]rediction error reports this information-theoretic free energy, which is mathematically constructed so as always to be greater than 'surprisal' (where this names the sub-personally computed implausibility of some sensory state given a model of the world […])." The relationship between free energy and surprisal then is that "[…] free energy is an upper bound on surprise, which means that if agents minimize free energy, they implicitly minimize surprise" (Friston 2010, p. 128). Suprisal, however, cannot be estimated directly by the system, because "there is an infinite number of ways in which the organism could seek to minimize surprise and it would be impossibly expensive to try them out" (Hohwy 2012, p. 3). The solution to this problem lies in implicitly minimizing surprisal (and its upper bound, i.e., free energy) by minimizing prediction error (cf. Hohwy 2013, p. 85, this collection, 3; see also Seth this collection, p. 6). It is exactly here where prediction

2 On a neuronal level of description, hierarchical generative models are assumed to be neuronally realized by multiple connections across low- and high-level cortical areas. Each level within the cortical hierarchy is connected to the next subordinate and supraordinate level, thereby ensuring effective inter-level message passing (cf. Hohwy 2013, pp. 67f). According to Clark (2013, p. 187), predictive generative models are implemented in "a kind of duplex architecture". This means that there are distinct neuronal units dedicated to the representation of predictions of environmental (or bodily) causes, so-called *representation units*, on the one hand, and those dedicated to the encoding of prediction error, so-called *error units*, on the other (cf. ibid.; Friston 2005, p. 829). To date, a detailed account of the concrete neuronal realization of these functionally distinct units of message-passing is still missing (cf. ibid.). However, it is hypothesized that representation units might correspond to superficial pyramidal cells, while error units might correspond to deep pyramidal cells (cf. Friston et al. 2012, p. 8; see also Clark 2013, pp. 187f).

error minimization avails itself as a tractable expression of more general life-sustaining mechanisms.

Prediction error minimization can be achieved in two distinct, yet complementary ways. The first of these is *perceptual inference*, which can be described as

> […] an iterative step-wise procedure where a hypothesis is chosen, and predictions are made, and then the hypothesis is revised in light of the prediction error, before new and hopefully better predictions are made on the basis of the revised hypothesis. (Hohwy 2013, p. 45)

That is, prediction errors are propagated up the hierarchy leading to an adjustment of the initial hypothesis, thereby achieving an approximation of the hypothesis generating the predictions and the actually given input. The adjustment of predictions and hypotheses in the face of fed-forward prediction error occurs at every level of the hierarchy until any prediction error is accommodated. This complex process comprising multiple levels is known as perception: "Perception thus involves 'explaining away' the driving (incoming) sensory signal by matching it with a cascade of predictions pitched at a variety of spatial and temporal scales" (Clark 2013, p. 187; see also Clark 2012, p. 762).

On Andy Clark's account of PP, one important consequence of this is that the traditional distinction between perception and cognition becomes blurred. It is replaced by a reconceptualization of perceptual and cognitive processes as a continuous employment of the same prediction error minimizing mechanism on multiple scales:

> All this makes the lines between perception and cognition fuzzy, perhaps even vanishing. In place of any real distinction between perception and belief we now get variable differences in the mixture of top-down and bottom-up influence, and differences of temporal and spatial scale in the internal models that are making predictions. Top-level (more 'cognitive') models intuitively correspond to increasingly abstract conceptions of the world, and these tend to capture or depend upon regularities at larger temporal and spatial scales. Lower-level (more 'perceptual') ones capture or depend upon the kinds of scale and detail most strongly associated with specific kinds of perceptual contact. (Clark 2013, p. 190)

Consequently, processes typically associated with perception or cognition can only be distinguished by considering the temporal and spatial resolution of the instantiation of PP mechanisms and the levels at which model revision ensues, respectively. This relationship between perception and cognition becomes important once we consider how enculturated cognition has been rendered possible on both phylogenetic and ontogenetic time scales. For it helps specify how evolutionary continuity could have been rendered possible in the first place. The evolutionary development of perception and cognition (and, as we shall see, of action too) may have proceeded from more perceptual generative models present in many other animals to more cognitive generative models exclusively realized in humans. This is in line with Roepstorff's (2013, p. 45) observation that "[t]he underlying neural models are basically species-unspecific, and the empirical cases move back and forth between many different model systems." Referring to this observation, Clark (this collection, p. 14) emphasizes that "[t]he basic elements of the predictive processing story, as Roepstorff (2013, p. 45) correctly notes, may be found in many types of organism and model-system." Thus, while certain (lower-level) model parameters and processing stages of prediction error minimization are shared by many organisms, there certainly are specific (higher-level) processing routines that are shared only by enculturated human organisms in a certain cognitive niche.

Furthermore, the idea that perception and cognition are continuous is relevant for considerations of the ontogenetic development of enculturated cognitive functions. This is because it anchors higher-order cognitive operations in more basic perceptual processes and thus allows for a fine-grained description of a certain developmental trajectory leading to cognitive transformation. Bearing in mind the hierarchical structure of generative models, another interesting consequence of the PP style approach to perception and cognition is that lower (i.e., more perceptual) levels of the generative model influence higher (i.e., more cognitive) levels by means of fed-forward prediction error. Vice versa, higher levels of the hierarchical generative model influence lower levels by means of fed-backward predictions (cf. Hohwy 2013, p. 73). This will become more important when we explore how reading acquisition can be described as an ongoing enculturating process of prediction error minimization.

Perceptual inference is only one way of minimizing prediction error. The second is *active inference*, where "[…] the agent will selectively sample the sensory input it expects" (Friston 2010, p. 129). The idea is that the system can minimize prediction error by bringing about the states of affairs (i.e., the environmental hidden causes) that are predicted by a certain hypothesis. This is achieved by performing any type of bodily movements, including eye movements, that make the selected prediction come true. The predictions at play in active inference are *counterfactual*, because

> […] they say how sensory input *would* change if the system *were* to act in a certain way. Given that things are not actually that way, prediction error is induced, which can be minimized by acting in the prescribed way. (Hohwy 2013, p. 82; italics in original; see also Clark this collection, p. 6; Friston et al. 2012, p. 2)

Accordingly, in active inference the selected prediction is held constant and leads to bodily activities that minimize prediction error by altering the sensory input such that it confirms the prediction. Therefore, active inference is of crucial importance for prediction error minimization, "[…] since it provides the only way (once a good world model is in place and aptly activated) to actually alter the sensory signal so as to reduce sensory prediction error" (Clark 2013, p. 202).

This suggests that perceptual and active inference, or perception and bodily action for that matter, mutually influence each other, thereby minimizing prediction errors and optimizing hypotheses generating ever new predictions. However, perceptual and active inference have a "different direction of fit" (Hohwy 2013, p. 178; see also Hohwy this collection, p. 13; Clark this collection, p. 7).[3] This is because in perceptual inference, predictions are aligned to the sensory input, while active inference is a matter of aligning the sensory input to the predictions. It follows "[…] that to optimally engage in prediction error minimization, we need to engage in perceptual inference and active inference in a complementary manner" (Hohwy 2013, p. 91). Since both perceptual and active inference are aimed at minimizing prediction error and optimizing generative models, "[p]erception and action […] emerge as two sides of a single computational coin" (Clark 2012, p. 760).

As emphasized earlier, perception and cognition are deeply related to the extent that both phenomena are the result of the same underlying functional and neuronal mechanisms. By extension, action is also deeply intertwined with cognition. This follows from the assumptions that 1. perception and cognition are continuous and 2. perception and action are subject to the same principles of prediction error minimization. As Seth (this collection, p. 5) puts it, both ways of prediction error minimization "[…] unfold continuously and simultaneously, underlining a deep continuity between perception and action […]." Yet, perceptual and active inference fulfil distinct functional roles in their ongoing attempt to minimize prediction error. This becomes even more obvious once we take the free energy principle into account: "The free energy principle […] does not posit any fundamental difference between perception and action. Both fall out of different reorganizations of the principle and come about mainly as different direc-

3 The notion of two functions having "a different direction of fit" originates in J. L. Austin's (1953, p. 234) speech act theory and in G. E. M. Anscombe's (1963, p. 56) example illustrating how words and states of affairs can relate to each other. I would like to thank Thomas Metzinger for pointing out the philosophical history of this notion.

tions of fit for prediction error minimization […]" (Hohwy this collection, p. 13). Active inference plays a crucial role in cognition (understood as prediction error minimization comprising many higher-level predictions), for it helps minimize prediction error throughout the cortical hierarchy by bringing about the states of affairs in the environment that are predicted on higher levels. Therefore, on Clark's (2013, p. 187) account, which he dubs *action-oriented predictive processing*, prediction error minimization "[…] depicts perception, cognition and action as profoundly unified and, in important respects, continuous."

PP accounts of human perception, action, and cognition distinguish between first-order and second-order statistics. In contrast to first-order statistics, which amount to minimizing prediction error by means of perceptual and active inference, second-order statistics are concerned with estimating the *precision* of prediction error. In second-order statistics, the influence of fed-forward prediction error on higher levels of the hierarchical generative model is dependent upon its estimated precision. Neuronally, the estimation of precision is captured in terms of increasing or decreasing the *synaptic gain* of specific error units (cf. Feldman & Friston 2010, p. 2). That is, "[t]he more precision that is expected the more the gain on the prediction error in question, and the more it gets to influence hypothesis revision" (Hohwy 2013, p. 66; see also Friston 2010, p. 132). Conversely, if the precision is expected to be poor on the basis of second-order statistics, the synaptic gain on the error unit is inhibited such that the prediction on the supraordinate level is strengthened (cf. ibid., p. 123). It has been proposed that precision estimation is equivalent to attention. This means that "attention is nothing but optimization of precision expectations in hierarchical predictive coding" (Hohwy 2013, p. 70; see also Feldman & Friston 2010, p. 2). For current purposes, it is sufficient to focus in the main on first-order statistics. However, it is important to bear in mind the crucial modulatory role precision estimation plays in prediction error minimization.

2.3 Combining cognitive integration and predictive processing

To what extent is it feasible to describe the mechanisms underlying cognitively integrated processes and enculturated cognition in terms of prediction error minimization? After having summarized CI and the core ideas of the PP framework I will argue in this section that there are many aspects of the CI approach that can be enriched by making a crucial assumption, namely that PP can account for many components constituting cognitive practices on at least functional and neuronal levels of description.

First, a major conceptual consequence of PP is that perception, action, and cognition are both continuous and unified, if this approach proves correct. This is because they follow the same principles of prediction error minimization, yet are characterized by important functional differences. This kind of complementarity fits neatly with the *hybrid mind thesis* defended by CI. Recall that the hybrid mind thesis claims that cognitive processes are constituted by both neuronal and extracranial bodily components. By taking prediction error minimization into account, this claim can be cashed out by assuming that the neuronal components are equal to perceptual inferences at multiple levels of the cortical hierarchy, while the bodily components are mechanistically realized by active inferences. The hybrid mind thesis emphasizes the indispensable, close and flexible coordination of neuronal and bodily components responsible for the completion of a cognitive task. The PP framework, or so I shall argue, provides the resources for a careful description of the underlying mechanisms at play. It does so by depicting human organisms as being constantly engaged in prediction error minimization by optimizing hypotheses in the course of perceptual inference and by changing the stimulus array in the course of active inference.

A second advantage of the prediction error minimization framework is that it helps cash out the *manipulation thesis*. This thesis, recall, states that "the manipulation of external vehicles [is] a prerequisite for higher cognition and embodied engagement [is] a precondition for these manipulative abilities" (Menary 2010, p. 232). In terms of the PP framework, bodily manipulation can be understood as an instance of active inference occurring in specific contexts. That is, in order to complete a certain cognitive task, the system changes its sensory input by altering certain components of its cognitive niche. This becomes even more obvious once we take into account that embodied activity is also a means of increasing confidence in sensory input by optimizing its precision. As suggested by Hohwy (this collection, p. 6), "expected precision drives action such that sensory sampling is guided by hypotheses that the system expects will generate precise prediction error." Applied to an organism's interaction with its socio-culturally shaped environment, Hohwy (2013, p. 238) argues "[…] that many of the ways we interact with the world in technical and cultural aspects can be characterized by attempts to make the link between the sensory input and the causes more precise (or less uncertain)." However, bodily manipulation is more than just a contributing factor to prediction error minimization (and precision optimization). In order to acknowledge this, we need to take into account that bodily manipulations are a crucial component of the performance of cognitive practices. In the performance of a cognitive practice, the minimization of prediction error and the optimization of precision is not an end in itself. Rather, it serves to facilitate the completion of a certain cognitive task. Furthermore, the concrete bodily manipulations given in terms of active inference are subject to cognitive norms that constrain the ways in which human organisms interact with cultural resources, such as tokens of a representational writing system. That is to say that the performance of a cognitive practice is not an individualistic enterprise. Rather, in completing a cognitive task, the individual is deeply immersed into a socio-cultural context which is shared by many human organisms.

Third, it is the normative constraints on cognitive practices that render their performance efficient and, in many cases at least, successful. This is because compliance with these norms induces what Andy Clark (2013, p. 195) calls "path-based idiosyncrasies". That is, one of the reasons why the coordination of neuronal and bodily components in the manipulation of cultural resources is beneficial certainly is that it takes place in a normatively constrained "multi-generational development of stacked, complex 'designer environments' for thinking such as mathematics, reading, writing, structured discussion, and schooling" (ibid.). That is to say that the performance of cognitive practices in compliance with certain norms has the overall advantage of reducing cognitive effort, which can be captured as the minimization of overall prediction error and the optimization of precision on a sub-personal level of description. At the same time, however, cognitive practices themselves can be described, or so I shall argue, as having prediction error minimization as their underlying mechanism. This double role of cognitive practices, described in terms of prediction error minimization, can be fully appreciated once we consider the cognitive transformations brought about by the ongoing interaction with cultural resources.

Fourth, our cognitive capacities and the various ways we complete cognitive tasks are profoundly augmented by our neuronal and bodily engagements with the socio-culturally structured environment through ontogenesis (cf. Menary 2006, p. 341). Put differently, "cognitive transformations occur when the development of the cognitive capacities of an individual are sculpted by the cultural and social niche of that individual" (Menary this collection, p. 8). This niche includes mathematical symbol systems, representational writing systems, artifacts, and so forth. It is this immersion and, importantly, the scaffolding provided by other inhabitants of the cognitive niche that ideally lead to the transformation of neuronal and extracranial bodily components constituting cognitive processes, to enculturation that is. The PP framework, or so I shall argue, offers a highly promising account of learning that is most suitable for a sub-personal level description of cognitive transformation. On the construal of PP, learning flows naturally from the mechanism of prediction error minimization. For learning can generally be construed as a sub-personally realized strategy of optimizing models and hypotheses in the face of ever new prediction error: "Learning is then viewed as the continual updating of internal model parameters on the basis of degree of predictive success: models are updated until they can predict enough of the signal" (Hohwy 2011, p. 268). Broadly understood, 'learning' thus figures as an umbrella term referring to the ongoing activity of prediction error minimization and model optimization throughout the lifetime of a human organism. This is because potentially ever new and "surprisaling" sensory signals need to be "explained away" by perceptual and active inference. For current purposes, however, "learning" can also be understood in a rather narrow sense as the acquisition of a certain skill, which is also subject to prediction error minimization through perception, action, cognition, and the modulation of attention. It is the individual's socio-culturally structured environment that delivers new sensory signals helping optimize parameters of the generative model:

> But those training signals are now delivered as part of a complex developmental web that gradually comes to include all the complex regularities embodied in the web of statistical relations among the symbols and other forms of socio-cultural scaffolding in which we are immersed. We thus self-construct a kind of rolling 'cognitive niche' able to induce the acquisition of generative models whose reach and depth far exceeds their apparent base in simple forms of sensory contact with the world. (Clark 2013, p. 195)

However, complex skills that are targeted at the completion of cognitive tasks cannot be learned simply by being exposed to the right kind of "training signal" in the cognitive niche. What is additionally needed is engagement in activities that are scaffolded by inhabitants of that cognitive niche who have already achieved a sufficient degree of expertise. This is what Menary (this collection) calls "scaffolded learning". From the perspective of PP, this amounts to the strategy of exposing predictive systems to highly structured, systematically ordered patterns of sensory input in the cognitive niche. This, however, needs to be complemented by a fine-grained personal-level description of the kind of interactions between experts and novices that is needed in order to pass on the right set of cognitive norms. Furthermore, the kind of cognitive transformation at play here requires a description of the neuronal changes that are correlated with the acquisition of a certain cognitive practice. That is, we need a more fine-grained account of LDP and how it might be realized in the human cortex. From the perspective of the PP framework, one plausible conjecture at this point is that LDP can be captured in terms of *effective connectivity*. Effective connectivity reports the causal interaction of neuronal assemblies across multiple levels of the cortical hierarchy (and across different brain areas) as a result of attention in terms of precision estimation. This line of reasoning is implied by Clark (2013, p. 190) who argues that "[a]ttention [...] is simply one means by which certain error-unit responses are given increased weight, hence becoming more apt to drive learning and plasticity, and to engage in compensatory action." This last point is important, since it stresses that it is not only perceptual inference that drives learning and contributes to the improvement of generative models, but also active inference. However, this approach to the acquisition of action patterns in concert with an optimization of precision might raise the worry that learning is depicted here as being a rather internalistic, brain-bound affair. But once we acknowledge that it is the performance and ongoing improvement of embodied active inferences that play an indispensable functional role in the completion of cognitive tasks, it becomes obvious that this worry is not warranted. For it is the efficient interaction of neuronal and extracranial bodily components (i.e., perceptual and active inferences in terms of PP) that results from learning and the efficient engagement of human organisms with their environment. Furthermore, LDP can now be considered in terms of the precision-weighted optimization of hypotheses throughout the cortical hierarchy and the ever new patterns of effective connectivity, as new cognitive practices are acquired and successfully performed. The sub-personal description of cognitive transformation in terms of prediction error minimization also does justice to neuronal reuse as a guiding principle of the allocation of neuronal resources for phylogenetically recent cognitive functions such as arithmetic or reading.

From this, the following question arises: What is the actual relationship between CI and PP supposed to be and what is the scope of this theory synthesis? First of all, the position developed in this commentary is neutral with regards to metaphysical consequences that may or may not result from the idea that CI and PP can be integrated into a unified theoretical framework. Rather, this position has an instrumentalist flavour to the extent that it tries to answer the question by which means socio-culturally shaped target phenomena can be best investigated both conceptually and empirically. Thus, the combination of CI and PP is valid only to the extent that it displays great descriptive as well as predictive power and is supported by many results stemming from empirical research. As such, the new approach on offer here is contingent upon the current state of research in cognitive science. It is falsifiable by new empirical evidence or convincing conceptual considerations that directly speak against it. Furthermore, it sidesteps the concern that PP and the underlying free energy principle might be trivial because they can be applied to any target phenomenon by telling a "just-so story". This is because the combination of CI and PP is applied to specific domains, namely to classes of cognitive processes that count as cognitive practices, with reading being the paradigm example.[4] Thus the approach advocated can be seen as a modest contribution to the project aiming at a "[…] translation into more precise, constricted applications to various domains, where predictions can be quantified and just-so stories avoided" (Hohwy this collection, p. 14).

The idea that CI and PP can be combined can lead to different degrees of commitment.[5] First, I do not assume that CI *necessarily requires* PP. Hypothetically, it is conceivable that another theory of neuronal and bodily functioning might be more suited to cashing out cognitive practices and enculturation more convincingly and more extensively. To date, PP appears to be the best unifying framework that helps specify exhaustively the functional and neuronal contributions of bodily and neuronal sub-processes giving rise to cognitive practices and enculturation. This is because PP offers a fine-grained functional and neuronal description of perception, action, cognition, attention, and learning that does justice to the complex interactions stipulated by CI and the associated approach to enculturation.

Second, it could be assumed that CI and PP are merely compatible. This would mean that CI and PP were self-sufficient and co-existent theoretical frameworks whose claims and key assumptions do not necessarily contradict each other. This compatibility assumption is too weak for various reasons thar have been presented in this commentary so far. For it is the purpose of the theory synthesis sketched here to enrich and refine the notion of enculturation and the associated theses defended by CI. Furthermore, to the extent that PP directly speaks to complex cognitive phenomena and learning, it benefits from the effort of CI to do justice to the socio-culturally shaped context in which these phenomena can be developed. This is to say that CI and PP can be directly referred to each other in ways that I have started to illustrate in this section.

Finally, from this it follows that both frameworks are more than just compatible – they are *complementary*. Taken together, they provide us with complex and far-reaching conceptual tools for investigating complex cognitive phenomena that are shaped by the individual's immersion in its cognitive niche. Thus, the complementarity of CI and PP leads to a new integrative framework that I dub enculturated predictive processing (EPP).

2.4 Defending enculturated predictive processing

At first glance, the EPP framework might appear to be unwarranted. For prediction error

4 Thanks to Jennifer M. Windt for raising this point.

5 Thanks to an anonymous reviewer for helpful suggestions on this issue.

minimization could be construed as being a purely internalistic, brain-bound affair that does not leave any room for the idea that cognitive processes are constituted both by neuronal and extracranial bodily components that are normatively constrained, socially scaffolded, and deeply anchored in a socio-culturally structured environment.

First, consider a position that takes for granted that cognitive processes can be coherently described in terms of prediction error minimization, but which denies that cognitive processes are co-constituted by neuronal and bodily sub-processes operating on socio-cultural resources. Such a position is defended by Jakob Hohwy (2013, p. 240) who argues that "[…] many cases of situated and extended cognition begin to make sense as merely cases of the brain attempting to optimize its sensory input so it, as positioned over against the world, can better minimize error." In particular, according to his interpretation of the prediction error minimization framework, "[…] the mind remains secluded from the hidden causes of the world, even though we are ingenious in using culture and technology to allow us to bring these causes into sharper focus and thus facilitate how we infer to them." (ibid., p. 239)

For Hohwy, this directly follows from the causal relations holding between the predictive system and the environmental causes it constantly tries to infer. According to him (ibid., p. 228), this relation needs to be characterized as "direct" and "indirect" at the same time:

> […] the intuition that perception is indirect is captured by its reliance on priors and generative models to infer the hidden states of the world, and the intuition that perception is direct is captured by the way perceptual inference queries and is subsequently guided by the sensory input causally impinging on it.

Since the causal relation that holds between a predictive system comprised of inverted generative models and the world is partly indirect, so the argument goes, the system is in constant embodied interaction and direct contact with its environment only insofar as it tries to make the effects of hidden causes fit the predictions. This precludes the theoretical possibility of depicting prediction error minimizing systems as being situated, scaffolded, integrated, or extended.

However, this line of reasoning fails to acknowledge the conceptual necessity of emphasizing the functional role of embodied active inference in terms of its contribution to the minimization of prediction error and the optimization of predictions. For even if the causal relations holding between a predictive, generatively organized system and environmental causes are mediated by hypotheses, predictions, prediction errors and precision estimation as encoded in the cortical hierarchy, it does not follow that this system is just a passive receiver of sensory input that informs it about remote states in the environment. Similarly, it does not necessarily follow from the prediction error minimization framework that it "[…] creates a sensory blanket – the evidentiary boundary – that is permeable only in the sense that inferences can be made about the causes of sensory input hidden beyond the boundary", as Hohwy (2014, p. 7) claims. Rather, the predictive system is part of its socio-culturally structured environment and has many possibilities for bodily acting in that environment in order to facilitate its own cognitive processing routines. Considering embodied active inference, it turns out that the causal relation holding between embodied action (in terms of bodily manipulation) and changes of the set of available stimuli in the environment is as direct as any causal relation could be. This is because these changes are an immediate effect of these very prediction error-minimizing and precision-optimizing actions, which in turn contribute to the performance of cognitive tasks. Furthermore, we need to take into account that genuinely human cognitive processes occur in a culturally sculpted cognitive niche, which is characterized by mathematical symbol systems, representational writing systems, artifacts, and the like, and other human organisms with whom we interact. These cognitive resources have unique properties that render them particularly useful for the completion of cognitive tasks.[6] For example, consider the regularity of line

6 Thanks to Richard Menary for raising this important point in personal communication.

arrangements and the orderliness of succeeding letters in an alphabetic writing system. Once learned and automatized, following these normative principles facilitates several types of cognitive processing routines. That is to say that it is the socio-culturally shaped sensory input itself that has an important impact on the concrete realization of prediction error minimization. This cannot be accounted for if we assume that the predictive processing of cognitive resources is an internalistic, secluded endeavour.

Second, consider a line of reasoning that goes against the compatibility of CI with the prediction error minimization framework, that might be put forward by an integrationist. She might agree that we need a mechanistic description of the neuronal and bodily components which jointly constitute cognitive processes in the close interaction with socio-cultural resources. But she might continue to argue that the performance of cognitive practices is more than just the minimization of prediction error and the optimization of precision.[7] From the perspective of PP, it needs neither to be denied that human cognitive systems as a whole aim to fulfil cognitive purposes by completing cognitive tasks and that they do so by engaging in cognitive practices. Nor should it be rejected that cognitive practices are normatively constrained and that cognitive systems are deeply immersed in a socio-culturally structured environment, which in turn provides these very norms through scaffolding teaching. However, the important theoretical contribution made by the prediction error minimization framework is its providing of a sub-personal, mechanistic description of the underlying neuronal and bodily sub-processes that turns out to be parsimonious, conceptually coherent, and empirically plausible. In addition, PP also offers a description of the close interaction of the neuronal and bodily components constituting cognitive practices by offering a concise description of the ongoing, mutually constraining interplay of perceptual and active inferences. More generally, this section should have established that all important claims and assumptions made by CI in favour of cognitive practices, such as the hybridity, the transformative efficacy, and the enculturated nature of cognitive processes, can be supplemented and refined by taking the prediction error minimization framework into account.

The arguments in favour of the EPP framework directly speak to the current debate within philosophy of mind and philosophy of cognitive science about the relationship between the prediction error minimization framework and approaches to situated, distributed, integrated, or extended cognition. On the one hand, Jakob Hohwy (2013, 2014) denies on both methodological and metaphysical grounds that there is anything like these types of cognition from the perspective of prediction error minimization. According to him, this is because predictive systems have only indirect access to the world. Furthermore, there is "the sensory boundary between the brain and the world" which prohibits predictive systems from engaging in any variant of situated, distributed, integrated, or extended cognition including CI (Hohwy 2013, p. 240). On the other hand, Andy Clark (2013, p. 195) argues that the PP framework at least "[…] offers a standing invitation to evolutionary, situated, embodied, and distributed approaches to help 'fill in the explanatory gaps' while delivering a schematic but fundamental account of the complex and complementary roles of perception, action, attention, and environmental structuring." Once we take the arguments and considerations in favour of EPP into account we have reasons to think that EPP lends support to Clark's construal of the PP framework. This will become even more persuasive once we take empirical data and a paradigm case of EPP into account.

3 Reading acquisition: A case of enculturation

So far, I have argued that the notion of enculturation and key claims made by CI can be enriched by taking the PP framework into account. In particular, the hybridity, embodiedness, and transformative character of enculturated cognition can be mechanistically described in terms of prediction error minimization. How-

7 This consideration was put forward by Richard Menary in personal communication.

ever, cognitive practices cannot be fully reduced to prediction error minimization, since they have a normative dimension that needs to be investigated on a personal level of description.

This section serves to illustrate the validity of the line of reasoning put forward in this commentary. This will be done by showing that reading acquisition, understood as another case of enculturation next to mathematical cognition, can be fruitfully described from the perspective of EPP.

3.1 Scaffolded learning and the acquisition of cognitive norms

One crucial aspect of learning to perform a cognitive practice is the acquisition of the relevant cognitive norms, where this class of norms "govern[s] manipulations of external representations, which aim at completing cognitive tasks" (Menary 2010, p. 238). In the case of reading, these norms concern the recognition and identification of tokens of a representational writing system. In alphabetic writing systems, important cognitive norms are derived from the so-called *alphabetic principle*, where this principle amounts to the "mapping [of] written units onto a small set of elements – the phonemes of a language" (Rayner et al. 2001, p. 33; see also Snowling 2000, p. 87). Specifically, the correspondence of graphemes to phonemes puts culturally established, normative constraints on the ways in which individual letters (and combinations thereof) are related to phonological units. The normative scope of these correspondences is best illustrated by differences across languages and orthographies. As pointed out by Ziegler & Goswami (2006, p. 430), "[i]n some orthographies, one letter or letter cluster can have multiple pronunciations (e.g. English, Danish), whereas in others it is always pronounced in the same way (e.g. Greek, Italian, Spanish)."[8] This demonstrates that the degree of consistency or transparency of *grapheme-phoneme correspondences* is subject to arbitrary stipulations by a linguistic, literate community employing a specific orthographic system. These stipulations are normative insofar as they constrain the ways in which combinations of letters are pronounced and written words are correctly related to spoken words. The acquisition of this normative knowledge needs "explicit instruction in the alphabetic principle" (Rayner et al. 2001, p. 57).[9] It follows that learning these norms is socially structured and dependent upon the cooperation of experts with novices. This fits neatly with Menary's (2013, p. 361) following assumption:

> Manipulative norms and interpretative norms apply to inscriptions of a public representational system and are never simply dependent on an individual. Indeed, it is the individual who must come to be transformed by being part of the community of representational system users.

Acquiring knowledge about grapheme-phoneme correspondences, especially in an inconsistent orthography such as English, puts demands not only on the novice, but also on the teachers who assist her in learning these correspondences. For the teachers, being experts in reading, need to break down their automatic identification and recognition skills in order to be able to teach the norms underlying the relationship between graphemes and phonemes. As Sterelny (2012, p. 145) points out more generally, "[e]xpert performance is often rapid and fluent, without obvious components. Learning from such performance is difficult. It becomes much easier if the task is overtly decomposed into segments, each of which can be represented and practiced individually." In the present context, the most successful strategy of teaching grapheme-phoneme correspondence has turned out to be so-called *phonics instruction* (cf. Rayner et al. 2001, pp. 31f): "[…] teaching methods that make the alphabetic principle explicit result in greater success among children trying to master the reading skills than methods that do not make it explicit" (ibid., p. 34). This goes along with teaching novices that spoken language consists of phonemes. That is, children's reading acquisi-

8 This phenomenon is also known as orthographic depth. For a recent review, see Richlan (2014).

9 See also Dehaene (2010, p. 219), Dehaene (2011, p. 26), and Frith (1985, p. 307).

tion is dependent upon, or at least co-develops with *phonological awareness*, where this is understood as "[…] the ability to perceive and manipulate the sounds of spoken words" (Castles & Coltheart 2004, p. 78). The *metalinguistic awareness* that spoken language consists of phonemes must be explicitly acquired and allows the novice to learn that these units correspond to letters, or combinations thereof. It is still debated whether phonological awareness is a prerequisite for learning to read or whether it is co-emergent with basic letter decoding skills. However, as suggested by Castles & Coltheart (2004, p. 104), "[…] it may not be possible for phonemic awareness to be acquired at all in the absence of instruction on the links between phonemes and graphemes." Thus, it seems safe to assume that phonological awareness clearly facilitates the ability to relate graphemes to phonemes. There are other components of metalinguistic awareness that influence the successful application of norms governing alphabetic representational writing systems. Beginning readers are already proficient speakers of their native language and are able to fluently apply syntactic, semantic, and pragmatic norms in their everyday conversations. However, they are usually unable to explicitly represent that utterances are made up of sentences and that sentences are made up of combinations of words (cf. Frith 1985, p. 308; Rayner et al. 2001, p. 35). To novices, these basic properties must be made explicitly available in order to put those novices in the position to apply knowledge about them automatically and fluently at later stages of reading acquisition. Furthermore, novices need to be acquainted with the convention, which is fairly obvious to expert readers, that alphabetic writing systems are decoded from left to right and from the top to the bottom of a page. These basic personal-level components of the acquisition of reading skills provide the cognitive norms necessary for the development of reading understood as a cognitive practice. It is these norms that govern the successful manipulation of representational vehicles belonging to an alphabetic writing system that need to be established by social interaction between learners and teachers. Thus, becoming proficient in applying the alphabetic principle, getting to grips with phoneme-grapheme correspondences, and developing phonological and metalinguistic awareness are cases of scaffolded learning.

3.2 Reading acquisition and neuronal transformation

Next to scaffolded learning, another crucial aspect of cognitive transformation is LDP (cf. Menary 2013, p. 356, this collection, p. 8). Indeed, in the case of reading acquisition, there is unequivocal evidence pointing to "[…] plastic changes in brain function that result from the acquisition of skills" (Ansari 2012, p. 93). By the same token, Ben-Shachar et al. (2011, p. 2397) emphasize that "[…] culturally guided education couples with experience-dependent plasticity to shape both cortical processing and reading development." As Schlaggar & McCandliss (2007, p. 477) point out, the application of knowledge about grapheme-phoneme correspondences in novice readers "[…] implicates the formation of functional connections between visual object processing systems and systems involved in processing spoken language." The left ventral occipitotemporal (vOT) area appears to play a crucial role in establishing these connections.

As mentioned by Menary (this collection), there has been consensus on the contribution of the vOT area to a neuronal reading circuit. In a series of experiments, Stanislas Dehaene, Laurent Cohen and their colleagues have made the remarkable discovery that neuronal activation in one particular region of the left vOT area is reliably and significantly associated with visual word recognition in adult, non-pathological readers (Cohen & Dehaene 2004; Dehaene 2005, 2010; Dehaene & Cohen 2011; Dehaene et al. 2005; McCandliss et al. 2003; Vinckier et al. 2007). This region, especially the left ventral occipito-temporal sulcus next to the fusiform gyrus, frequently responds to visually presented words regardless of the size, case, and font in which they are made available (cf. Dehaene 2005, p. 143; McCandliss et al. 2003, p. 293). This consistent finding has led these researchers to call it the visual word form area (VWFA), since it crucially contributes to "[…] a critical process that groups the letters of a word together into an integrated perceptual unit (i.e. a 'visual word form')" (McCandliss et al. 2003, p. 293). However, it is debatable whether the left vOT area is almost exclusively dedicated to visual word recognition in expert readers, or whether this area serves several functions having to do with the (visual) identification of shapes more broadly construed (see Price & Devlin 2003, 2004, for a discussion). Nevertheless, the findings by Dehaene and his colleagues that the left vOT area plays a crucial role in the overall visual word recognition process is important and widely acknowledged, although the interpretations of its functional contribution differ.

An important motivation for research on the overall function of the left vOT area stems from considerations on the phylogenetic development of visual word recognition. Considering that writing systems were invented only approximately 5400 years ago, it is unlikely that the ability to read is the result of an evolutionary process (cf. Dehaene 2005, p. 134, 2010, p. 5; McCandliss et al. 2003, p. 293). In a nutshell, the crucial question is how visual word recognition is possible given "[…] that the human brain cannot have evolved a dedicated mechanism for reading" (Dehaene & Cohen 2011, p. 254). This is also referred to as the "reading paradox" (Dehaene 2010, p. 4). The solution to this paradox proposed by Dehaene and his colleagues is to assume "[…] that plastic neuronal changes occur in the context of strong constraints imposed by the prior evolution of the cortex" as a result of the human organism being exposed to tokens of a certain writing system (Dehaene & Cohen 2011, p. 254). Specifically, the idea is "[…] that writing evolved as a recycling of the ventral visual cortex's competence for extracting configurations of object contours" (ibid.). This view, which has been dubbed the *neuronal recycling hypothesis* (cf. Dehaene 2005, p. 150), suggests that existing neuronal functions associated with visual cognition are "recycled" for the phylogenetically recent, ontogenetically acquired capacity to recognize visually presented words (cf. Cohen & Dehaene 2004, p. 468; see also Menary 2014, p. 286). This "recycling" is in turn constrained by the overall evolved neuronal architecture and already existing processing mechanisms (cf. Dehaene 2010, pp. 146f). Thus, neuronal recycling is just a special type of neuronal reuse (see Anderson 2010, for a discussion). There are certain conditions that need to be met if a specific cortical area is to be 'recycled' for a phylogenetically recent cognitive function (see Menary 2014, p. 288). In the case of visual word recognition, the left vOT area is assumed to exert certain "functional biases" that make it most suitable for the recognition and identification of visually presented words: "(1) a preference for high-resolution foveal shapes; (2) sensitivity to line configurations; and (3) a tight proximity, and, presumably, strong reciprocal interconnection to spoken language representations in the lateral temporal lobe" (Dehaene & Cohen 2011, 256). These "functional biases", however, do not preclude that the left vOT area is still engaged in other cognitive processes such as object recognition in skilled adult readers (cf. Carreiras et al. 2014, p. 93; Dehaene & Cohen 2011, p. 257; Price & Devlin 2004, p. 478). Rather, it helps explain why this area is found to be well-equipped for contributing to the overall process of visual word recognition. However, the question arises what the contribution of the left vOT area to the overall visual word recognition process is supposed to make. According to Cathy Price's & Joseph Devlin's (2011) Interactive Account (IA), the contribution of the left vOT area can be best described and explained in terms of PP. In line with the general principles of the PP framework presented above, they generally hold the following assumption: "Within the hierarchy, the function of a region depends on its synthesis of bottom-up sensory inputs conveyed by forward connections and top-down predictions mediated by backward connections" (Price & Devlin 2011, p. 247). In other words, the suggested synthesis equals the prediction error that results from the discrepancy between top-down predictions and bottom-up sensory information. Applied to the patterns of neuronal activation associated with visual word recognition, this assumption is specified as follows:

> For reading, the sensory inputs are written words (or Braille in the tactile modality) and the predictions are based on prior association of visual or tactile inputs with phonology and semantics. In cognitive terms, vOT is therefore an interface between bottom-up sensory inputs and top-down predictions that call on non-visual stimulus attributes. (Price & Devlin 2011, p. 247)

Accordingly, the vOT area is supposed to be associated with a distinct level of the hierarchical generative model responsible for visual word recognition mediating between higher-level, language-related predictions and bottom-up visual information. It follows that "[…] the neural implementation of classical cognitive functions (e.g. orthography, semantics, phonology) is in distributed patterns of activity across hierarchical levels that are not fully dissociable from one another" (ibid., p. 249). Specifically, IA proposes a neuronal mechanism that is able to demonstrate how linguistic knowledge about phonology and semantics, encoded in top-down predictions, causally interacts with bottom-up information. This is because it is held that a prediction error is generated each time bottom-up information diverges from the associated top-down prediction. In turn, the resulting prediction error is associated with significant activation in the left vOT area. Empirical evidence supporting this approach to the functional contribution of the left vOT area to visual word recognition in expert readers is widely available (see, e.g., Bedo et al. 2014; Kherif et al. 2011; Kronbichler et al. 2004; Schurz et al. 2014; Twomey et al. 2011).

In reading acquisition, the left vOT area appears to be an equally important contributor to visual word recognition. According to Price & Devlin (2011, p. 248), the activation level of the vOT area develops in a non-linear fashion, as the proficiency in visual word recognition increases:

> In pre-literates, vOT activation is low because orthographic inputs do not trigger appropriate representations in phonological or semantic areas and therefore there are no top-down influences […]. In early stages of learning to read, vOT activation is high because-top-down predictions are engaged imprecisely and it takes longer for the system to suppress prediction errors and identify the word […]. In skilled readers, vOT activation declines because learning improves the predictions, which explain prediction error efficiently […].

That is, IA assumes that the level of activation within the left vOT area is dependent upon the general establishment and refinement of a generative model comprising both lower-level areas associated with visual processing and higher-level cortical areas associated with phonological and semantic knowledge. If this account turns out to be correct, the blurredness of the distinction between perception and cognition as suggested by Clark (2013) becomes vitally important. For it is the mutual interplay of lower-level processing stages (traditionally associated with visual processing) and higher-level processing stages (traditionally associated with phonological and semantic processing) that renders the successful acquisition of visual word recognition possible in the first place. Evidence in favour of IA comes from studies demonstrating that there is a significant increase of activation in this area as a result of exposure to visually presented words in beginning readers across different research paradigms and methodologies employing fMRI (e.g., Ben-Shachar et al. 2011; Gaillard et al. 2003; Olulade et al. 2013). Furthermore, two longitudinal ERP studies (Brem et al. 2010; Maurer et al. 2006) demonstrate that the left-lateralized occipito-temporal N1 effect, an effect associated with print sensitivity, does not develop in a linear fashion in the course of reading acquisition. Rather, Maurer et al.'s (2006, p. 756) comparison of their results obtained from their child participants with an adult control group indicates that "[i]nstead of a linear increase with more proficient reading, the development is strongly nonlinear: the N1 specialization peaks after learning to read in beginning readers and then decreases with further reading practice in adults following an inverted U-shaped developmental time-course." In this vein, Brem et al. (2010, p. 7942) interpret their results by suggesting that "[t]he emergence of print sensitivity in cortical areas during the acquisition of grapheme-phoneme correspondences is in line with the inverse U-shaped developmental trajectory of print sensitivity of the ERP N1, which peaks in beginning readers […]."

Another consequence of Price's & Devlin's (2011) PP account of reading acquisition is that the activation level within the vOT should be associated with the degree of accuracy of top-down predictions in the face of bottom-up signals. This is supported by various studies demonstrating that higher-level activations of cortical areas associated with language processing are also present in beginning readers. For example, Turkelhaub et al. (2003, p. 772) report that "[a]ctivity in the left ventral inferior frontal gyrus increased with reading ability and was related to both phonological awareness and phonological naming ability. […] Brain activity in the anterior middle temporal gyrus also increased with reading ability", where this area is associated with semantic processing. Similarly, Gaillard et al. (2003) report activation in the middle temporal gyrus, which is frequently associated with semantic processing in expert readers (e.g., Bedo et al. 2014, p. 2; Price & Mechelli 2005, p. 236; Vogel et al. 2013, p. 231; Vogel et al. 2014, p. 4). Furthermore, they report significant activation patterns in left IFG, which is associated with both phonological and semantic processing.

In the light of much empirical evidence in favour of Price's & Devlin's (2011) approach to the neuronal changes corresponding to reading acquisition, it seems safe to assume that it is empirically plausible and can account for many data derived from experiments in cognitive neuroscience. However, to what extent can this approach be conceptually enriched? Recall that learning a new skill such as reading is just a special case of overall prediction error minimization according to the PP framework. On this construal, learning to read means becoming increasingly efficient in predicting linguistic, visually presented input as a result of long-term exposure to types of this input and the optimization of hypotheses through perceptual inference. The careful instruction in relating graphemes to phonemes, phonological and metalinguistic awareness, and the normatively constrained alphabetic principle provides the environmental conditions for efficient and progressively more accurate prediction error minimization. The signals delivered by this highly structured learning environment are estimated as being precise, such that the synaptic gain on error units reporting the discrepancy between (still inaccurate) predictions and prediction error is high. As learning to read proceeds, the predictions become more accurate and the overall influence of prediction error shows a relative decrease. This line of reasoning is supported by Price's & Devlin's (2011, p. 248) following suggestion: "At the neural level, learning involves experience-dependent synaptic plasticity, which changes connection strengths and the efficiency of perceptual inference." Understood this way, LDP and the associated neuronal transformations can be understood as being realized by prediction error minimization in the context of scaffolded learning, which allows a beginning reader to become ever more efficient and successful in this particular cognitive practice.

3.3 Reading acquisition and bodily transformation

Starting from the hybrid mind thesis defended by CI, which states that certain cognitive processes are constituted by both neuronal and extracranial bodily sub-processes, it seems natural to assume that reading acquisition also is associated with the transformation of bodily sub-processes. That is, in the course of enculturation it is the enactment of bodily manipulation that is transformed in addition to the neuronal changes occurring as a result of LDP. In terms of PP, this assumption leads to the suggestion that it is not only perceptual inferences that are causally relevant for learning described in terms of prediction error minimization, but also active inferences that allow for ever more efficient subpersonally employed strategies for "explaining away" incoming sensory input. Recall that eye movements are just a special case of active inference (see e.g., Friston et al. 2012). Their functional contribution to prediction error minimization becomes vitally important for a complete account of visual word recognition and its acquisition. This is because visual word recognition, in both novices and experts, is rendered possible by the coordination of perceptual and active inference. From the perspective of CI, the idea here is that the ways in which an individual bodily manipulates a certain cognitive resource is importantly improved in the course of cognitive transformation. Applied to reading acquisition, this leads to the assumption that specific eye movement patterns become more efficient as a result of reading instruction and iterate exposure to a certain type of cognitive resource (say, sentences printed on a piece of paper).

Recently, it has become possible to investigate eye movements in beginning readers by employing eye-tracking methodologies. Converging evidence suggests that beginning readers make more fixations (i.e., acquisition of visual information in the absence of oculomotor activities), saccades (i.e., oculomotor activities), and regressions (i.e., backward saccades), and exhibit longer fixation durations and smaller saccade amplitudes than proficient and expert readers (cf. Joseph et al. 2013, p. 3; Rayner et al. 2001, p. 46). More specifically, these tendencies are assessed in a longitudinal eye-tracking study reported by Huestegge et al. (2009). They measured eye movements during an oral reading task in second and fourth graders of a German primary school and additionally assessed overall reading skills and oculomotor behaviour beyond reading (cf. Huestegge et al. 2009, p. 2949). Their results indicate that the fourth graders, in comparison to the second graders, show a decrease of fixation duration, gaze duration, total reading time, refixations, and saccadic amplitudes (cf. ibid., p. 2956). Huestegge et al. (2009, p. 2958) attest that the younger, less proficient readers show a "[…] refixation strategy, with initial saccade landing positions located closer to word beginnings." Similarly to Huestegge et al. (2009), Seassau et al. (2013) report a longitudinal study comparing the performance of 6- to 11-year-old children in a reading task and a visual task. In line with the empirical evidence already mentioned, their results indicate that "[w]ith age, children's reading capabilities improve and they learn to read by making larger progressive saccades, fewer regressive saccades and shorter fixations […]" (Seassau et al. 2013, p. 6). Furthermore, it is demonstrated that the eye movement patterns employed in reading and in visual search diverge with increasing reading proficiency (cf. ibid., p. 9).

An explanation of these results in terms of PP is straightforward. In beginning readers, the predictions initiating active inference occurring in a highly-structured linguistic environment are inaccurate, such that the generation and execution of eye movements in terms of active inference is not as efficient as it is in the case of expert readers. By the same token, the inaccuracy of the currently selected prediction makes it necessary to sample the visually available linguistic environment more thoroughly, explaining the "refixation strategy" and the execution of comparatively more saccades. As reading skills improve, resulting from increasingly efficient prediction error minimization through perceptual inference as already suggested, the accuracy of predictions becomes increasingly optimal, therefore allowing for more efficient active inference. More efficient active inference, in turn, allows for more efficient perceptual inference, since both types of inference mutually influence each other. This line of reasoning is supported by Huestegge et al.'s (2009, p. 2957) claim informed by the results of their study "[…] that only linguistic, not oculomotor skills were the driving force behind the acquisition of normal oral reading skills." Thus, the increase in efficiency of eye movements in beginning readers does not result from an increase in oculomotor capabilities *per se*, but works in tandem with higher-level linguistic knowledge encoded in predictions, which are associated with representations in higher-order cortical areas. As a result, the improvement of active inference in the course of reading acquisition works in tandem with the improvement of perceptual inference. This highlights that learning to read does not only result in neuronal, but also in bodily transformations. As such, the optimization of eye movements in the course of reading acquisition highlights the importance of bodily manipulation in the efficient enactment of reading understood as a cognitive practice. This also means to suggest that a complete account of enculturation should not only pay attention to scaffolded learning and LDP, but also to the developmental trajectory of bodily manipulation.

4 Concluding remarks

This commentary on Richard Menary's paper *Mathematical Cognition: A Case of Enculturation* started from the assumption that the general outline of enculturation and the associated claims made by CI provide important conceptual tools for the description of ontogenetically acquired, socio-culturally shaped cognitive processing routines. However, I have argued that the idea of enculturation and its most important aspects, namely cognitive transformation and scaffolded learning, need to be enriched by providing a detailed functional and neuronal description on a sub-personal level of description. In addition, it needs to be born in mind that enculturation is rendered possible by normative constraints developed by a large group of individuals sharing the same cognitive niche. To this end, I have suggested that the notion of enculturation and its associated constitutive aspects can be complemented in important ways by taking the PP framework into account. The result is what I call enculturated predictive processing. Thus, the PP framework is capable of providing the conceptual resources necessary for a thorough description of the mechanistic underpinnings of cognitive practices and their acquisition. Lending further support to this line of reasoning, I have dealt with reading acquisition as a paradigmatic case of enculturated predictive processing. This should have been sufficient to establish that the CI framework is well-suited for a conceptually coherent description of the interaction between brain, body, and environmental cognitive resources. However, it needs to be supplemented by a sub-personal level description in terms of prediction error minimization in order to be able to specify the neuronal and functional underpinnings of the hybrid mind thesis, the bodily manipulation thesis, and the transformation thesis as defended by CI. At the same time, the approach to reading acquisition put forward in this commentary suggests that a vast array of empirical findings from cognitive neuroscience and cognitive psychology can be unified for the first time by interpreting them from the new perspective of enculturated predictive processing. Thus, I submit that we can only appreciate the cognitive assets rendered possible by our socio-culturally structured environment once we account for the enabling conditions of sophisticated, neuronally and bodily realized cognitive processes such as mathematical cognition and reading. These conditions include socio-culturally established ways of learning and teaching, LDP, and the ability to adapt action patterns to the needs and requirements of a certain cognitive task. My overall claim is that we need the EPP framework to be able to approach the entire spectrum of these factors, whose complex interplay ultimately leads to truly enculturated cognition.

Acknowledgements

The author wishes to thank the Barbara Wengeler Foundation for its generous financial support. In addition, she is indebted to Thomas Metzinger, Jennifer M. Windt, and an anonymous reviewer for their helpful feedback on earlier versions of this commentary.

References

Anderson, M. L. (2010). Neural reuse: A fundamental organizational principle of the brain. *Behavioral and Brain Sciences*, *33* (04), 245-266. 10.1017/S0140525X10000853

Anderson, M. L. & Finlay, B. L. (2014). Allocating structure to function: The strong links between neuroplasticity and natural selection. *Frontiers in Human Neuroscience*, *7*. 10.3389/fnhum.2013.00918

Ansari, D. (2012). Culture and education: New frontiers in brain plasticity. *Trends in Cognitive Sciences*, *16* (2), 93-95. 10.1016/j.tics.2011.11.016

Anscombe, G. E. M. (1963). *Intention (2nd ed.)*. Cambridge, Mass: Harvard University Press.

Austin, J. L. (1953). How to talk: Some simple ways. *Proceedings of the Aristotelian Society* (53), 227-246.

Bedo, N., Ribary, U., Ward, L. M. & Valdes-Sosa, P. A. (2014). Fast dynamics of cortical functional and effective connectivity during word reading. *PLoS ONE*, *9* (2), e88940-e88940. 10.1371/journal.pone.0088940

Ben-Shachar, M., Dougherty, R. F., Deutsch, G. K. & Wandell, B. A. (2011). The development of cortical sensitivity to visual word forms. *Journal of Cognitive Neuroscience*, *23* (9), 2387-2399. 10.1162/jocn.2011.21615

Brem, S., Bach, S., Kucian, K., Guttorm, T. K., Martin, E. & Lyytinen, H. (2010). Brain sensitivity to print emerges when children learn letter-speech sound correspondences. *Proceedings of the National Academy of Sciences*, *107* (17), 7939-7944. 10.1073/pnas.0904402107

Carreiras, M., Armstrong, B. C., Perea, M. & Frost, R. (2014). The what, when, where, and how of visual word recognition. *Trends in Cognitive Sciences*, *18* (2), 90-98. 10.1016/j.tics.2013.11.005

Castles, A. & Coltheart, M. (2004). Is there a causal link from phonological awareness to success in learning to read? *Cognition*, *91* (1), 77-111. 10.1016/S0010-0277(03)00164-1

Clark, A. (2012). Dreaming the whole cat: Generative models, predictive processing, and the enactivist conception of perceptual experience. *Mind*, *121* (483), 753-771. 10.1093/mind/fzs106

——— (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, *36* (03), 181-204. 10.1017/S0140525X12000477

——— (2015). Embodied prediction. In T. Metzinger & J. M. Windt (Eds.) *Open MIND*. Frankfurt a. M., GER: MIND Group.

Cohen, L. & Dehaene, S. (2004). Specialization within the ventral stream: The case for the visual word form area. *NeuroImage*, *22* (1), 466-476. 10.1016/j.neuroimage.2003.12.049

Dehaene, S. (2005). Evolution of human cortical circuits for reading and arithmetic: The “neuronal recycling” hypothesis. In S. Dehaene, J.-R. Duhamel, M. D. Hauser & G. Rizzolatti (Eds.) *From monkey brain to human brain. A Fyssen foundation symposium* (pp. 133-157). Cambridge, MA: MIT Press.

——— (2010). *Reading in the brain: The new science of how we read.* New York, NY: Penguin Books.

——— (2011). The massive impact of literacy on the brain and its consequences for education. *Human neuroplasticity and education. Pontifical Academy of Sciences*, *117*, 19-32.

Dehaene, S. & Cohen, L. (2011). The unique role of the visual word form area in reading. *Trends in Cognitive Sciences*, *15* (6), 254-262. 10.1016/j.tics.2011.04.003

Dehaene, S., Cohen, L., Sigman, M. & Vinckier, F. (2005). The neural code for written words: A proposal. *Trends in Cognitive Sciences*, *9* (7), 335-341. 10.1016/j.tics.2005.05.004

Feldman, H. & Friston, K. J. (2010). Attention, uncertainty, and free-energy. *Frontiers in Human Neuroscience*, *4*. 10.3389/fnhum.2010.00215

Friston, K. (2005). A theory of cortical responses. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *360* (1456), 815-836. 10.1098/rstb.2005.1622

——— (2010). The free-energy principle: A unified brain theory? *Nature Reviews Neuroscience*, *11* (2), 127-138. 10.1038/nrn2787

Friston, K., Adams, R. A., Perrinet, L. & Breakspear, M. (2012). Perceptions as hypotheses: Saccades as Experiments. *Frontiers in Psychology*, *3*. 10.3389/fpsyg.2012.00151

Frith, U. (1985). Beneath the surface of developmental dyslexia. In K. E. Patterson, J. C. Marshall & M. Coltheart (Eds.) *Surface dyslexia. Neuropsychological and cognitive studies of phonological reading* (pp. 301-330). Hillsdale, NJ: Erlbaum.

Gaillard, W., Balsamo, L., Ibrahim, Z., Sachs, B. & Xu, B. (2003). fMRI identifies regional specialization of neural networks for reading in young children. *Neurology*, *60* (1), 94-100. 10.1212/WNL.60.1.94

Hohwy, J. (2011). Phenomenal variability and introspective reliability. *Mind & Language*, *26* (3), 261-286. 10.1111/j.1468-0017.2011.01418.x

——— (2012). Attention and conscious perception in the hypothesis testing brain. *Frontiers in Psychology*, *3*. 10.3389/fpsyg.2012.00096

——— (2013). *The predictive mind.* Oxford, UK: Oxford University Press.

——— (2014). The self-evidencing brain. *Noûs*. 10.1111/nous.12062

——— (2015). The neural organ explains the mind. In T. Metzinger & J. M. Windt (Eds.) *Open MIND*. Frankfurt a. M., GER: MIND Group.

Huestegge, L., Radach, R., Corbic, D. & Huestegge, S. M. (2009). Oculomotor and linguistic determinants of reading development: A longitudinal study. *Vision Research*, *49* (24), 2948-2959. 10.1016/j.visres.2009.09.012

Joseph, H. S. S. L., Liversedge, S. P. & Paterson, K. (2013). Children's and adults' on-line processing of syntactically ambiguous sentences during reading. *PLoS ONE*, *8* (1), e54141-e54141. 10.1371/journal.pone.0054141

Kherif, F., Josse, G. & Price, C. J. (2011). Automatic top-down processing explains common left occipito-temporal responses to visual words and objects. *Cerebral Cortex*, *21* (1), 103-114. 10.1093/cercor/bhq063

Kronbichler, M., Hutzler, F., Wimmer, H., Mair, A., Staffen, W. & Ladurner, G. (2004). The visual word form area and the frequency with which words are encountered: Evidence from a parametric fMRI study. *NeuroImage*, *21* (3), 946-953. 10.1016/j.neuroimage.2003.10.021

Maurer, U., Brem, S., Kranz, F., Bucher, K., Benz, R., Halder, P., Steinhausen, H.-C. & Brandeis, D. (2006). Coarse neural tuning for print peaks when children learn to read. *NeuroImage*, *33* (2), 749-758. 10.1016/j.neuroimage.2006.06.025

McCandliss, B. D., Cohen, L. & Dehaene, S. (2003). The visual word form area: Expertise for reading in the fusiform gyrus. *Trends in Cognitive Sciences*, *7* (7), 293-299. 10.1016/S1364-6613(03)00134-7

Menary, R. (2006). Attacking the bounds of cognition. *Philosophical Psychology*, *19* (3), 329-344. 10.1080/09515080600690557

——— (2007a). *Cognitive integration: Mind and cognition unbounded.* New York, NY: Palgrave Macmillan.

——— (2007b). Writing as thinking. *Language Sciences*, *29* (5), 621-632. 10.1016/j.langsci.2007.01.005

——— (2010). Cognitive integration and the extended mind. In R. Menary (Ed.) *The extended mind* (pp. 227-243). Cambridge, MA: MIT Press.

——— (2013). The enculturated hand. In Z. Radman (Ed.) *The hand, an organ of the mind. What the manual tells the mental* (pp. 561-593). Cambridge, MA: MIT Press.

——— (2014). Neural plasticity, neuronal recycling and niche construction. *Mind & Language*, *29* (3), 286-303. 10.1111/mila.12051

——— (2015). Mathematical cognition: A case of enculturation. In T. Metzinger & J. M. Windt (Eds.) *Open MIND*. Frankfurt a. M., GER: MIND Group.

Menary, R. & Kirchhoff, M. (2014). Cognitive transformations and extended expertise. *Educational Philosophy and Theory*, 1-14. 10.1080/00131857.2013.779209

Olulade, O. A., Flowers, D. L., Napoliello, E. M. & Eden, G. F. (2013). Developmental differences for word processing in the ventral stream. *Brain and Language*, *125* (2), 134-145. 10.1016/j.bandl.2012.04.003

Price, C. J. & Devlin, J. T. (2003). The myth of the visual word form area. *NeuroImage*, *19* (3), 473-481. 10.1016/S1053-8119(03)00084-3

——— (2004). The pro and cons of labelling a left occipitotemporal region "the visual word form area". *NeuroImage*, *22* (1), 477-479.

——— (2011). The interactive account of ventral occipito-temporal contributions to reading. *Trends in Cognitive Sciences*, *15* (6), 246-253. 10.1016/j.tics.2011.04.001

Price, C. J. & Mechelli, A. (2005). Reading and reading disturbance. *Current Opinion in Neurobiology*, *15* (2), 231-238. 10.1016/j.conb.2005.03.003

Rayner, K., Foorman, B. R., Perfetti, C. A., Pesetsky, D. & Seidenberg, M. S. (2001). How psychological science informs the teaching of reading. *Psychological Science in the Public Interest*, *2* (2), 31-74. 10.1111/1529-1006.00004

Richlan, F. (2014). Functional neuroanatomy of developmental dyslexia: The role of orthographic depth. *Frontiers in Human Neuroscience*, *8*. 10.3389/fnhum.2014.00347

Roepstorff, A. (2013). Interactively human: Sharing time, constructing materiality. *Behavioral and Brain Sciences*, *36* (03), 224-225. 10.1017/S0140525X12002427

Rowlands, M. (1999). *The body in mind: Understanding cognitive processes, Cambridge studies in philosophy.* Cambridge, UK: Cambridge University Press.

Schlaggar, B. L. & McCandliss, B. D. (2007). Development of neural systems for reading. *Annual Review of Neuroscience*, *30* (1), 475-503. 10.1146/annurev.neuro.28.061604.135645

Schurz, M., Kronbichler, M., Crone, J., Richlan, F., Klackl, J. & Wimmer, H. (2014). Top-down and bottom-up influences on the left ventral occipito-temporal cortex during visual word recognition: An analysis of effective connectivity. *Human Brain Mapping*, *35* (4), 1668-1680. 10.1002/hbm.22281

Seassau, M., Bucci, M.-P. & Paterson, K. (2013). Reading and visual search: A developmental study in normal children. *PLoS ONE*, *8* (7), e70261-e70261. 10.1371/journal.pone.0070261

Seth, A. K. (2015). The cybernetic Bayesian brain: From interoceptive inference to sensorimotor contingencies. In T. Metzinger & J. M. Windt (Eds.) *Open MIND*. Frankfurt a. M., GER: MIND Group.

Snowling, M. J. (2000). *Dyslexia.* Malden, MA: Blackwell Publishers.

Sterelny, K. (2012). *The evolved apprentice: How evolution made humans unique,The Jean Nicod lectures: Vol. 2012.* Cambridge, MA: The MIT Press.

Turkeltaub, P. E., Gareau, L., Flowers, D. L., Zeffiro, T. A. & Eden, G. F. (2003). Development of neural mechanisms for reading. *Nature Neuroscience*, *6* (7), 767-773. 10.1038/nn1065

Twomey, T., Kawabata Duncan, K. J., Price, C. J. & Devlin, J. T. (2011). Top-down modulation of ventral occipito-temporal responses during visual word recognition. *NeuroImage*, *55* (3), 1242-1251. 10.1016/j.neuroimage.2011.01.001

Vinckier, F., Dehaene, S., Jobert, A., Dubus, J. P., Sigman, M. & Cohen, L. (2007). Hierarchical coding of letter strings in the ventral stream: Dissecting the inner organization of the visual word-form system. *Neuron*, *55* (1), 143-156. 10.1016/j.neuron.2007.05.031

Vogel, A. C., Church, J. A., Power, J. D., Miezin, F. M., Petersen, S. E. & Schlaggar, B. L. (2013). Functional network architecture of reading-related regions across development. *Brain and Language*, *125* (2), 231-243. 10.1016/j.bandl.2012.12.016

Vogel, A. C., Petersen, S. E. & Schlaggar, B. L. (2014). The VWFA: It's not just for words anymore. *Frontiers in Human Neuroscience*, *8*. 10.3389/fnhum.2014.00088

Ziegler, J. C. & Goswami, U. (2006). Becoming literate in different languages: similar problems, different solutions. *Developmental Science*, *9* (5), 429-436. 10.1111/j.1467-7687.2006.00509.x

What? Now. Predictive Coding and Enculturation

A Reply to Regina E. Fabry

Richard Menary

Regina Fabry has proposed an intriguing marriage of enculturated cognition and predictive processing. I raise some questions for whether this marriage will work and warn against expecting too much from the predictive processing framework. Furthermore I argue that the predictive processes at a sub-personal level cannot be driving the innovations at a social level that lead to enculturated cognitive systems, like those explored in my target paper.

Keywords
Active inference | Cognitive integration | Enculturation | Learning driven plasticity | Mathematical cognition | Perceptual inference | Predictive processing | Reading

Author

Richard Menary
richard.menary@mq.edu.au
Macquarie University
Sydney, NSW, Australia

Commentator

Regina E. Fabry
fabry@students.uni-mainz.de
Johannes Gutenberg-Universität
Mainz, Germany

Editors

Thomas Metzinger
metzinger@uni-mainz.de
Johannes Gutenberg-Universität
Mainz, Germany

Jennifer M. Windt
jennifer.windt@monash.edu
Monash University
Melbourne, Australia

1 Introduction: What? Now.

I'd like to thank Regina Fabry for her excellent and detailed response to my paper. She articulates an important account of reading acquisition as a process of enculturation and describes how a Cognitive Integration/Enculturated Cognition (henceforth CI/ENC) account can be combined with a predictive processing account of neural processing. She shows, in impressive detail, how CI/ENC can benefit from Predictive Processing (henceforth PP), primarily as a way of explaining the neural-level details of processes that conspire with bodily interactions with the local environment to complete cognitive tasks. Since Fabry's response suggests an important way of cashing out some of the details of an enculturated approach, I would like to take this opportunity to look at some of the potential pitfalls in the proposed Enculturated Predictive Processing style (henceforth EPP). Primarily I want to focus on the differences in explanatory emphasis between CI/ENC and PP, especially where CI/ENC proposes the importance of the population-level effects of normative patterned practices (henceforth NPP), such as mathematical practices.

PP is all about predictions, happening in the here-and-now[1]; however CI/ENC occurs at different levels and over much longer time-scales. It turns out that this difference is important, because if the brain is engaged in predictive error minimization (as sub-personal processing) in the here-and-now, then it cannot be driving the innovation of new NPP over many generations. This is because the pressures driving those innovations are found at the social, or populational, level[2], not at the level of neural processing where 'what?' is answered in the now.

I also raise several issues concerning the nature of the PP project, particularly whether, as a theory of general brain architecture, all processing can be cashed out in terms of predictive processes. I'm also sceptical about Fabry's claim that PP can provide the "mechanistic underpinnings of the acquisition of cognitive practices" (Fabry this volume, p. 3) on its own, without help from what I call learning-driven plasticity (LDP) and neural redeployment. Finally, I comment on the promising research path down which Fabry is headed.

In the first section I remind the reader of some of the leading ideas of the CI/ENC framework, highlighting, in particular, the different levels of explanation and how this matters to the proposed marriage of ENC-PP. In the second section I raise several problems for the PP approach in general and for the ENC-PP approach in particular. My main concerns are to push away from an 'isolated brain' interpretation of PP and to place EPP within a much broader context of explanation.

2 CI and enculturation

As I point out in my contribution to this volume, cognitive integration should be understood as a thesis about the enculturation of human cognition. It is a thesis about how phylogenetically earlier forms of cognition are built upon by more recent cultural innovations (e. g., systems of symbolic representation). This results in a multi-layered system with heterogeneous components, dynamically interwoven into a co-operative of processes and states an integrated cognitive system (henceforth ICS). The co-ordination dynamics of the system are, at least in part, understood in terms of the physical dynamics of brain–body–niche interactions in real-time; however, they are also to be understood in terms of NPP that govern and determine those interactions (over time). NPP operate at both social/population levels and individual, even sub-personal, levels. They originate as patterns of activity spread out over a population of agents; consequently they should be understood primarily as public systems of activity and/or representation that are susceptible to innovative alteration, expansion, and even contraction over time. They are transmitted horizontally across generational groups and vertically from one generation to the [3]next. At the individual level they are acquired, most often by learning and training, and they manifest themselves as changes in the ways in which individuals think, but also the ways that they act (intentionally) and the ways in which they interact with other members of their social group(s) and the local environment. NPP, therefore, operate at different levels (groups and individuals) and over different time-scales (intergenerationally and in the here-and-now).

Given this, it is clear that What? Now[4] processes that reduce prediction errors on their own could not drive the innovation of NPP; nor could they determine the properties of NPP on their own. Less obviously, I would argue, they do not drive the acquisition of NPP, because scaffolded learning requires both a physically and temporally-structured learning environment and the capacity for functional changes to cortical circuitry to be driven by the structured learning environment. The mechanism of acquisition includes both neural and environmental processes working in concert and over long periods of ontogenetic time. What? Now processes may help us to understand the here-and-now

1 I mean predictions on incoming sensory input relevant to immediate action in the environment.

2 I think that these levels are real. There is a level of entire populations, social groups, individual organisms and there is a level of individual brains. Cognition takes place within and across (at least the final three) levels.

3 See my target article for examples.

4 Predictions on sensory input in the here-and-now.

processes by which we enact NPP; they may even tell us something about the neural mechanisms for learning and plasticity; but we should be wary of making prediction and error minimization the driving factors behind the why and how of enculturation.

Fabry's commentary focuses on the neural level, functioning in real-time, where the primary aim is to give a mechanistic account of how cognitive capacities can be transformed by learning and training in rich socio-cultural niches. Rather than looking at the origin of ICS in cultural inheritance, phenotypic plasticity, and learning driven plasticity, Fabry argues that a version of the PP framework can provide the neural mechanisms by which ICS are (partly) constructed. My contribution to this volume focused primarily on the origin of ICS in the recent cultural evolution of NPP and then explored how mathematical practices could be learnt and how this process of learning could drive functional changes to circuitry in the brain. Consequently, the CI/ENC framework pursues the phylogenetic and ontogenetic basis of the larger brain–body–niche nexus. What, though, of the neural mechanisms of transformation?

I don't agree with Fabry's starting premise that CI/ENC lacks a mechanism of transformation: the mechanism of transformation is learning-driven plasticity (LDP) with neural redeployment in a scaffolded learning environment. The fundamental plasticity of the brain explains the nature of neural transformations and why the brain is open to scaffolded learning driven by the environment. (E)PP does not have the resources to explain redeployment (this is a theme I take up in the next section). Why would it, since PP is not a framework for explaining redeployment. It might be the case that PP fits with a certain conception of scaffolded learning such as path-dependent learning, but I have yet to see a thorough working-through of the details and it's not clear to me that all scaffolded learning should be reduced to a predictive form of path-dependent learning.

Fabry claims that a dynamical systems approach to integration "does not spell out the mutual influence that neuronal and extra-cranial bodily components have over each other" (2015, p. 3). The EPP approach is supposed to fill in the details here. However, I suspect that this judgement is made a little too quickly, because the dynamical systems description of brain–body–niche interactions is in one sense a higher-level description of those interactions. The dynamical interactions are described as being part of a larger system comprising brain, body, and niche. We can zoom in and focus upon the dynamics of brain or body, but we shouldn't confuse the dynamics of the brain for the dynamics of the overall system. I have highlighted and outlined the neural dynamics required for enculturation in a number of places. For example, in the account of body schema dynamics and in the case of NPP for symbolic cognition, I have outlined the case for dual component transformations (e. g., Menary 2007, pp. 78–83; 2010; 2013 and 2014). Lets take these two cases in order.

In a now famous series of studies, Maravita & Iriki (2004) studied the bimodal interparietal neurons in trained Japanese macaque brains. These neurons respond both to tactile stimulation on the hand (tactile receptive field) and visual stimuli in the same vicinity as the tactile receptive field (the visual receptive field). The visual receptive field was centred on the hand following it through space. When macaques where trained to use a rake to pull food towards them on a table, the observation that struck Maravita and Iriki was that when the macaques used the rake the receptive fields of the bimodal neurons extended along the axis of the rake, including its head. Iriki's interpretation of this is that "either the rake was being assimilated into the image of the hand or, alternatively, the image of the hand was extending to incorporate the tool" (Iriki & Sakura 2008, p. 2230). The extension of the body schema (receptive field) to include the tool happened only during active holding; it reduced to just the hand during inactivity. The interesting result of these experiments is that the existing body schema has the latent capacity to extend to incorporate the tool. LDP can be cashed out in terms of functional changes as the result of scaffolded learning even in the case of macaques, let alone the notoriously plastic brains of humans.

Functional changes can be cashed out in terms of neural redeployment and cortical connectivity. Returning to the case of mathematical cognition, inherited systems for numerosity are evolutionary endowments; we can be reasonably sure of this because they are constant across individuals and cultures and they are shared with other species. The numerosity systems are "quick and dirty"; they are approximate and continuous, not discrete and digital. By contrast, discrete mathematical operations exhibit cultural and individual variation; there is a big difference between Roman numerals and Arabic numerals. They are subject to verbal instruction (they actually depend on language); one must learn to count, whereas one does not learn to subitise. Mathematics depends on cultural norms of reasoning (mathematical norms). The ability to perform exact mathematical calculations depends on the public system of representation and its governing norms. We learn the interpretative practices and manipulative practices as a part of a pattern of practices within a mathematics community, and these practices transform what we can do. They are constitutive of our exact calculative abilities. Mathematical practices get under our skins by transforming the way that our existing neural circuitry functions.

The relationship between the evolutionarily earlier system and the recent development of public mathematical systems, norms, and symbols comes down to the redeployment of the cortical territories that are dedicated to evolutionarily older functions by novel cultural artefacts (e. g., representations, tools). The transformation results in new connections between the frontal lobe for number-word recognition and association, the temporal lobe for the visual recognition of number form, and the parietal lobe for the approximate recognition of magnitudes across both left and right hemispheres (Dehaene 1997).

The deeply transformative power of our learning histories in the cognitive niche relates to the development of our capacities for understanding symbolic representations and for physically manipulating inscriptions in public space. In learning to understand symbols, the first transformation involves our sensorimotor abilities for creating and manipulating inscriptions (the transformation of the body schema). This is something we learn to do on the page and in the context of a learning environment, in public space, before we do it in our heads. Our capacities to think have been transformed, but in this instance they are capacities to manipulate inscriptions in public space.

It looks like PP can provide models of some of the fundamental processing principles at work at the sub-personal neural level, but it is not obvious that it would replace LDP and neural redeployment in the mechanism of transformation. However, Fabry may be right and PP may add another string to the bow of our understanding of how the brain exhibits the plasticity required for cognitive transformation. In that case it provides extra explanatory depth to the account of enculturation, but only as part of a much broader explanatory framework.

3 Some worries for enculturated predictive coding

Fabry provides a persuasive case for how PP could provide the neural underpinnings of enculturation. In this section, however, I will raise some problems for the proposed marriage of CI/ENC and PP. The main issues I will address are as follows:

1. The incompatibility of the isolated brain interpretation (Hohwy 2013) and the active inference interpretation (Clark 2013) of PP.
2. The attempt to explain all cognitive processing in terms of prediction error.
3. The redeployment of neural circuitry as not being explained by PP.
4. The role of NPP as not being explained by the reduction of prediction error.

1. Isolating the brain

If CI/ENC has one central commitment, it is that we should not think of cognition as isolated from the environment. And yet this is exactly how we ought to understand the predictive brain, according to a prominent interpretation of the PP framework. Whenever the PP framework is introduced, it is almost always introduced in the following way: "Accounts of PP generally assume that human perception, action, and cognition are realized by *Bayesian probabilistic generative models* implemented in the human brain. Since *the human brain does not have immediate access to the environmental causes of sensory effects, it has to infer the most probable state of affairs in the environment* giving rise to sensory data" (Fabry 2015, p. 4; my emphasis). The two main motivations for the PP framework are that the brain is isolated from the environment and must make a best guess as to what it is perceiving, and that this kind of probabilistic inference-making results in internal (neurally realized) models of the environment. Putting aside the probabilistic nature of the inferences, this just is old-fashioned individualism. There is a perceptual interface to an environment of hidden variables; the internal system creates internal models (representations) of those hidden environmental variables, which then causally produce behaviour. The internal states must predict the external variables via sensory input, but they have no direct access to the causal ancestry of the sensory input. This form of individualism is used as an explanation for why models and predictions are required: "Because the brain is isolated behind the veil of sensory input, it is then advantageous for it to devise ways of optimizing the information channel from the world to the senses" (Hohwy 2013, p. 238). Hohwy describes the mind–world relation as "fragile" because of the isolation of the brain, and this is why active inference is required.

The saving grace of the PP framework, from the perspective of CI/ENC, is active inference. In Clark's version of PP active inference and cultural props help to minimize prediction errors (Clark 2013); and because of this, there is a deep continuity between mind and world mediated by active inference and the cultural scaffolding of our local niche. Curiously, Hohwy agrees with Clark's interpretation, but at a cost. Hohwy agrees that active inference and the cultural scaffolding of the environment help to change sensory input so as to minimize prediction error, but also "by increasing the precision of the sensory input" (Hohwy 2013, p. 238). According to Hohwy, the primary role of PP is perceptual inference; as a matter of "second order statistics" active inference helps to optimise sensory input so that perceptual inference is less error-prone.

Note the cost. First, active inference and cultural scaffolding is relegated to the secondary role of reducing prediction error for the primary cognitive job of perceptual inference, which is carried out wholly by matching statistical models to sensory input in the brain. Second, Hohwy shows that this interpretation of active inference should be understood against the background of the isolated brain. "The key point I am aiming at here is that this is a picture that accentuates the indirect, skull-bound nature of the prediction error minimization mechanism" (Hohwy 2013, p. 238). Organizing and structuring our environments makes sense if the mind–world relation is fragile in the way that Hohwy presents it, and also because this structuring makes perceptual inference more reliable. I take it that Fabry and Clark would deny this interpretation of the role of active inference and cultural scaffolding. Indeed, Fabry denies Hohwy's 'isolationist' interpretation in her commentary.

However, Fabry does so by playing up the roles of NPP, which go far beyond prediction minimization: "Furthermore, we need to take into account that genuinely human cognitive processes occur in a culturally sculpted cognitive niche. [...] These cognitive resources have unique properties that render them particularly useful for the completion of cognitive tasks" (Fabry 2015, p. 12). She also nods to the subpersonal, mechanistic role of PP in the entire brain–body–niche nexus: "[T]he important theoretical contribution made by the prediction error minimization framework is its providing of a sub-personal, mechanistic description of the underlying neuronal and bodily sub-processes" (Fabry 2015, p. 13). It is therefore not clear to me that PP does anything more than provide the functional details of *some* of the neural processing in the brain–body–niche nexus. It certainly should not be taken to provide a comprehensive account of what cognition is and why there is cultural scaffolding, or what its interesting cognitive properties are.[5] It is to these issues that I shall now turn.

2. Everything is predicted

One of the main concerns with the PP approach is that it is used both to try to explain all of cognition and as an explanation of why there is cultural scaffolding. We've already seen a brief hint of this in Hohwy, Clark, and Fabry's work above.[6] The first worry can be found in the expression of PP as originating in the free energy principle:

> The free-energy considered here represents a bound on the surprise inherent in any exchange with the environment, under expectations encoded by its state or configuration. A system can minimise free energy by changing its configuration to change the way it samples the environment, or to change its expectations. These changes correspond to action and perception, respectively, and lead to an adaptive exchange with the environment that is characteristic of biological systems. This treatment implies that the system's state and structure encode an implicit and probabilistic model of the environment. (Friston & Stephan 2007, p. 417)

PP is primarily a model of the way in which top-down processing 'predicts' bottom up sensory input and which samples the environment to change its expectations. These correspond to perception and action respectively[7]. However, it seems odd to build a cognitive theory on the basis of the prediction of sensory signals. This is because much of cognition is not about sensory signal prediction; nor about actions as sampling the environment. Indeed much of cognition isn't about 'prediction' at all. So whilst I agree that at least part of the mechanisms of cognition can be fruitfully modelled by PP, not all of them will be. In enculturated systems, the really important work is being done by the processing governed by normative patterned practices whose properties are understood primarily at the social or populational level. I agree that at the individual level, the mechanisms of ICS can partly be explained by PP, but the main explanatory work will not be a matter of predictions of sensory input[8].

The examples from Landy & Goldstone (2007) may be partly explained by prediction errors, but again this only makes sense in the context of sensorimotor processing governed by mathematical norms. If the norms function as priors in the system, then this might help explain the errors made by the test subjects.

3. Phenotypic plasticity and neural redeployment

PP can't explain the redeployment of neural circuitry to new cognitive functions. And it is not supposed to, since this isn't the job it was designed to do. However, this is a considerable weakness if PP is supposed to be the primary mechanism of enculturation. I've already canvassed the reasons why in section 1.

4. NPP and prediction error minimization

Enculturated PP plays a role in the multilayered and interwoven ICS, but it neither determines nor implements the entire system. My argument in this response has been that the dynamics of ICS are not determined by the predictive processing of parts of the system: if any-

5 CI/ENC provides just these motivations and details. Clark himself proposes that the PP framework "offers a standing invitation to evolutionary, situated, embodied, and distributed approaches to help 'fill in the explanatory gaps' while delivering a schematic but fundamental account of the complex and complementary roles of perception, action, attention, and environmental structuring" (Clark 2013, p. 195).

6 See also their contributions to this volume.

7 There are also theories of attention based upon PP, but I won't address those here.

8 Thomas Metzinger has raised an interesting question for me here: whether there is continuity between the levels? My argument has been that there is continuity between the levels, but this continuity is made possible by NPP's, LDP and neural redeployment. PP explains how we make perceptual inferences about the environment and it might explain something about the hierarchical organisation of neural architecture. However, it should be seen as playing *a* role in the organisation and enculturation of the brain, not the *only* role.

thing PP is enslaved to the processing needs of the entire enculturated system. The PP framework takes perceptual inference as its primary mode of processing, which is the top-down matching of predictions to sensory input. However, it is not obvious that this is the right model for all cognitive processing, since it is not obvious that all cognitive processing is just a matter of predictions about sensory input, nor a hierarchically organised system which minimises prediction errors.

For example Hohwy (2013, p. 238) argues that "many of the ways we interact with the world in technical and cultural aspects can be characterized by attempts to make the link between the sensory input and the causes more precise (or less uncertain)." This would be a very impoverished account of the evolution of public systems of representation. Public systems of representation did not simply evolve to "make the link between the sensory input and the causes more precise (or less uncertain)"; this would be to ignore the social pressures that would have caused representational innovation.[9] It might be true that the history of the refinement of notation has something to do with making input more easy to 'predict'; however, this would not be an *ultimate* explanation for why there are notations in the first place, nor how they function in our cognitive lives. It *might* be a *proximal* explanation of the neural mechanisms for the processing of notations and as such, it might explain some of the causal conditions that explain how notations have developed, but it doesn't explain the conditions under which notations evolved. For further reasons why see section 3.4 of my target article, on evolutionary novelty and uniqueness (this volume).

For example, the idea that the brain predicts the product of two numerals makes sense, and the surprise at a product too distant from the operands lends further credence. Remember the example from section 4.1 of my target article (this volume) : 34 + 47 = 268. However, it is not obvious that predictions will help with the second example: 34 x 47 = 1598. What is required in this instance is the serial working through of the multiplication according to an algorithm. Furthermore, this is not simply a case of sensory predictions: when it comes to recognising the numerals on the page in front of you, PP can explain top-down predictions about sensory input, but that is not at all the same thing as the working through of a mathematical problem. So mathematical cognition could not, it seems to me, be reducible to error minimization.

9 I take it that Hohwy is claiming that cultural representations function so as to make perceptual inferences more precise. This would be another way of reducing socio-cultural phenomena to a role that is complementary to the brain, with the processing needs of the brain dictating the evolutionary path that culture must take. The externalist perspective takes it that there are social and cultural pressures that require cognitive innovations (sometimes even new phenotypes).

4 Conclusion: Where now?

Despite some of my concerns about how the PP framework can be interpreted and its relation to the CI/ENC framework, I think that Fabry's account of the enculturation of reading using a hybrid of CI and EPP is really compelling. This leads me to think that an EPP account might by workable for other cases, such as mathematical cognition. Having said this, the division of labour between PP and evolutionary accounts of the origin of NPP and ICS must be in place. The role of scaffolded learning and neural redeployment should not be replaced by error minimization processes. The 'isolationist' reading of PP should be resisted, and a more situated cognition friendly approach embraced. PP is a sub-personal account of neural processes that fits within a larger account of the brain–body–niche nexus. If one embraces CI/ENC then there's more to the mind than What? Now.

Acknowledgements

Thanks to Thomas Metzinger and Jenny Windt for comments and to Regina Fabry for her excellent commentary.

References

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, *36* (3), 181-204. 10.1017/S0140525X12000477

Dehaene, S. (1997). *The number sense: How the mind creates mathematics.* London, UK: Penguin.

Fabry, R. E. (2015). Enriching the Notion of Enculturation: Cognitive Integration, Predictive Processing, and the Case of Reading Acquisition - A Commentary on Richard Menary. In T. Metzinger & J. M. Windt (Eds.) *Open MIND*. Frankfurt a. M., GER: MIND Group.

Friston, K. & Stephan, K. E. (2007). Free energy and the brain. *Synthese*, *159*, 417-458. 10.1007/s11229-007-9237-y

Hohwy, J. (2013). *The predictive mind.* Oxford, UK: Oxford University Press.

Iriki, A. & Sakura, O. (2008). The neuroscience of primate intellectual evolution: natural selection and passive intentional niche construction. *Philosophical Transactions of the Royal Society B*, *363*, 2229-2241. 10.1098/rstb.2008.2274

Landy, D. & Goldstone, R. L. (2007). How abstract is symbolic thought? *Journal of Experimental Psychology*, *33* (4), 720-733. 0.1037/0278-7393.33.4.720

Maravita, A. & Iriki, A. (2004). Tools for the body (schema). *Trends in Cognitive Sciences*, *8*, 79-86. 10.1016/j.tics.2003.12.008

Menary, R. (2007). *Cognitive integration: Mind and cognition unbounded.* London, UK: Palgrave Macmillan.

——— (2010). Dimensions of mind. *Phenomenology and the Cognitive Sciences*, *9*, 561-578. 10.1007/s11097-010-9186-7

——— (2013). The enculturated hand. In Z. Radman (Ed.) *The hand, an organ of the mind. What the manual tells the mental* (pp. 593-561). Cambridge, MA: MIT Press.

——— (2014). Neuronal recycling,neural plasticity and niche construction. *Mind and Language*, *29* (3), 286-303. 10.1111/mila.12051

——— (2015). Mathematical Cognition - A Case of Enculturation. In T. Metzinger & J. M. Windt (Eds.) *Open MIND*. Frankfurt a. M., GER: MIND Group.