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 neur-

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:* 25(C). Frankfurt am Main: MIND Group. doi: 10.15502/9783958571143

onal 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 occipitotemporal 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