Structure and function in the predictive brain

Marco Facchin, Ph.D. [**corresponding author**]
Research Fellow
Antwerp University
Centre for Philosophical Psychology
Prinstraat 13, Antwerp 2000
marco.facchin.marco.facchin@gmail.com
https://orcid.org/0000-0001-5753-987

Marco Viola, Ph.D.
Junior Assistant Professor
Roma tre University
Department of Philosophy,
Communication and Performing Arts
Via Ostiense 234, Rome 00144
marco.viola@uniroma3.it
https://orcid.org/0000-0001-7970-5406

Abstract:

Predictive processing is an ambitious neurocomputational framework, offering an unified explanation of all cognitive processes in terms of a single computational operation, namely prediction error minimization. Whilst this ambitious unificatory claim has been thoroughly analyzed, less attention has been paid to what predictive processing entails for structure-function mappings in cognitive neuroscience. We argue that, taken at face value, predictive processing entails an *all-to-one* structure-function mapping, wherein each individual neural structure is assigned the same function, namely minimizing prediction error. Such a structure-function mapping, we show, is highly problematic. For, barring few, rare occasions, such a structure-function mapping fails to play the predictive, explanatory and heuristic roles structure-function mappings are expected to play in cognitive neuroscience. Worse still, it offers a picture of the brain that we know is wrong. For, it depicts the brain as an *equipotential* organ; an organ wherein structural differences do not correspond to any appreciable functional difference, and wherein each component can substitute for any other component without causing any loss or degradation of functionality. Somewhat ironically, the very neuroscientific roots of predictive processing motivate a form of skepticism concerning the framework's most ambitious unificatory claims. Do these problems force us to abandon predictive processing? Not necessarily. For, once the assumption that all cognition can be accounted for *exclusively* in terms of prediction error minimization is relaxed, the problems we diagnosed lose their bite.

***Keywords**:* Predictive processing, cognitive ontology, structure-function mapping, equipotentialism

Declarations:

Founding: MFis founded by the FWO grant "Towards a globally non-representational theory of the mind" (grant number 1202824N)

Conflict of Interests/competing interests: The author declares no conflict of interests.

Availability of data and Material: Not applicable.

Code availability: Not applicable.

Authors' contribution: MF and MV contributed equally to the project. MF gets to be the first author only because he's the one who came up with the idea.

Acknowledgements: We wish to thank Luke Kersten, Beate Krickel and Marcin Milkowski for having read and commented upon previous iterations of this essay. Many thanks also to Lisa Feldman Barrett for a long and inspiring discussion of the paper. A very early version of the works has also been presented at the 2024 Rudolf Carnap Lectures (Ruhr University Bochum). We thank the audience for their stimulating questions.

Structure and function in the predictive brain

Abstract:

Predictive processing is an ambitious neurocomputational framework, offering an unified explanation of all cognitive processes in terms of a single computational operation, namely prediction error minimization. Whilst this ambitious unificatory claim has been thoroughly analyzed, less attention has been paid to what predictive processing entails for structure-function mappings in cognitive neuroscience. We argue that, taken at face value, predictive processing entails an *all-to-one* structure-function mapping, wherein each individual neural structure is assigned the same function, namely minimizing prediction error. Such a structure-function mapping, we show, is highly problematic. For, barring few, rare occasions, such a structure-function mapping fails to play the predictive, explanatory and heuristic roles structure-function mappings are expected to play in cognitive neuroscience. Worse still, it offers a picture of the brain that we know is wrong. For, it depicts the brain as an *equipotential* organ; an organ wherein structural differences do not correspond to any appreciable functional difference, and wherein each component can substitute for any other component without causing any loss or degradation of functionality. Somewhat ironically, the very neuroscientific roots of predictive processing motivate a form of skepticism concerning the framework's most ambitious unificatory claims. Do these problems force us to abandon predictive processing? Not necessarily. For, once the assumption that all cognition can be accounted for *exclusively* in terms of prediction error minimization is relaxed, the problems we diagnosed lose their bite.

***Keywords**:* Predictive processing, cognitive ontology, structure-function mapping, equipotentialism

1 - Introduction

Cognitive neuroscience attempts to provide a scientific understanding of how the mind works and how it is implemented in its neural substrate. A popular strategy to this end is that of *functional localization*: the mapping of specific mental operations, often referred to as "cognitive functions", onto specific neural structures (McCaffrey 2023).

These mappings are supposed to play numerous relevant roles in cognitive neuroscience. First, they should play a *predictive* role. Knowing the functions of a neural structure should enable us to estimate which cognitive tasks recruit ("activate") it, and to use patterns of neural activations to estimate which task is being executed. Secondly, structure-function mappings should play an *explanatory* role, as structure and functions should be mutually illuminating. The need of playing a certain function should explain *why* a structure boasts its specific organization; and its organization should in turn explain *how* the function is actually executed. Lastly, *heuristically*, structure-function mapping should suggest novel ways to probe neurocognitive functions and chart our neural architecture.

Structure-function mappings are thus central for cognitive neuroscience to reach its goals. In the early days of the discipline, neuroimagers settled for establishing one-to-one correspondences between folksy defined cognitive functions and individual, anatomically localized neural areas (e.g. "language tasks increase activation in *Broca's area*"). Yet, progressively cognitive neuroscientists replaced such "neo-phrenological", one-to-one correspondences with more nuanced mapping holding between *non-folksily-defined* functions onto *networks* of co-activated areas (cf. Poldrack 2010; Sporns 2014; De Brigard & Gessell *forthcoming*).

But this has not been the only relevant development in cognitive neuroscience. In the last decade or so, a neuro-computational framework known as Predictive Processing (PP) has set sail not "just" to explain mind and cognition in its entirety (see Hohwy 2013; Clark 2013, 2016; Spratling 2016), but also to provide a global account of the brain and its organization (Friston 2009, 2010, 2013). Importantly, PP casts our neurocognitive system as an hierarchy of processing units all devoted to the same task, that of minimizing an intracerebral signal known as prediction error. Different cognitive processes such as perception, action, attention and emotion, are thus reduced to different upshot of a single underlying neurocomputational process.

Such a view of the brain has received enormous philosophical and scientific attention, mostly focused on its promise to explain different cognitive phenomena (e.g. Hohwy 2020; Sun & Firestone 2020; Clark 2023) and to unify the cognitive sciences under a single banner (Colombo & Wright 2017; Litwin & Miłkowski 2020; Poth 2022). Less attention has been paid to what PP entails for cognitive neuroscience specifically. Our paper aims to partially fill in this gap, examining what PP entails for functional localization. We will show that PP entails an all-to-one structure function mapping, which assigns the same function to all neural structures, namely the function of minimizing prediction error. Such a structure-function mapping, we will argue, is ill-suited to play the predictive, explanatory and heuristic roles structure-function mappings are expected to play. In fact, it ends up depicting the brain as an equipotential organ - something we know the brain isn't. This doesn't, however, mean that we should simply abandon PP. For, PP can avoid these problems simply by abandoning its ambitions of providing a complete explanation of brain and cognition, thereby making space for alternative functional ascriptions.

Our paper is structured as follows. §§ 2 and 3 lay down the groundwork for our analysis. The former introduces the epistemic roles structure-function mappings are expected to play in the cognitive neurosciences; the latter introduces PP. §4 then capitalizes on this material to argue that PP entails an all-to-one structure-function mapping, wherein all neural structures are assigned one and the same cognitive function, which is the function of minimizing prediction error. This structure-function mapping, however, fails to play the relevant epistemic roles structure-function mappings are supposed to play in cognitive neuroscience. Worse still, the all-to-one structure function mapping PP entails casts the brain as an equipotential organ - something that, given our current knowledge, is almost certainly false. §5 examines and addresses some possible objections. Lastly, §6 offers a short conclusion, suggesting that, to avoid the problems we have highlighted, PP should give up its most ambitious explanatory aims.

2 - Structure-function mappings in cognitive neuroscience

Before explaining what roles we think structure-function mappings must play in cognitive neuroscience, let us clarify what we mean by "structure" and "function". We use "structure" broadly, to designate all cognitively relevant "bits" of the brain, be them areas, regions, fascicules or networks. Similarly, "function" designates cognitive constructs such as those employed by cognitive neuroscientists, e.g. "cognitive control", "working memory", "binding features" or "minimizing prediction errors". Notice that we are interested exclusively in cognitive functions. Other functions (e.g., monitoring the oxygen level in the blood, cf. Haueis 2018) are certainly relevant to understanding the brain, but we won't deal with them here. Our talk of structure-function mapping should not be read as an implicit commitment to a naive localizationalism, or to the idea that each and every neural structure plays a single-context invariant, functional role. To put it bluntly, structure-function mappings, in the sense we are using the term, don't need to be one-to-one: they can be many-to-many and context sensitive - and indeed, given our current knowledge, there is good reason to expect structure-function mappings to be many-to-many (Anderson et al. 2013)and/or context-sensitive in most cases (Burnston 2016). Still, in all cases, structure-function mappings involve the assignment of particular function(s) to individual neural structure(s).

As anticipated, structure-function mappings should play at least three relevant epistemic roles in cognitive neuroscience. First, they should play a predictive role, enabling us to estimate the probability of a pattern of activation, given that the subject is performing a task (see Price & Friston 2005). And vice versa: based on certain patterns of activation, we should be able to estimate which cognitive functions have likely been recruited and the task that is likely being performed. Further, by observing that certain structures are damaged, we should be able to expect certain functions to be impaired. An ideal neurocognitive theory should allow us to make such estimations with the highest possible accuracy. Take, as an example of this predictive role, the renown fusiform face area (FFA). It has been that some portion of the right temporal cortex is selectively activated during the observation of faces (Kanwisher et al. 1997).

Furthermore, its stimulation leads to hallucinate faces (Schalk et al. 2017). Based on these findings, a clinical neurologist can reasonably hypothesize that a neurological patient with an impaired ability to recognise faces, i.e., suffering from acquired prosopagnosia, presents some lesions in the FFA. It is also reasonable to expect that novel face-related tasks will activate the FFA; and so, that subjects whose FFA is activated are dealing with face-related tasks.

Structure-function mappings should also play an explanatory role: the way neural structure are internally organized and connected to each other should explain how neural functions are performed (cf. Craver 2007; Bechtel 2008). Neural structures differ in many respects (e.g. in their histology, connectivity, and receptive fields), and these differences presumably correspond to different functional roles. So, it is natural to appeal to the former to explain how the latter are carried out - for example, to suggest that the relevance of FFA for face perception may be driven also by an internal organization privileging face-like patterns and its connectivity bias, both downstream (as it receives projections from early visual areas most receptive of curvilinear stimuli represented at the center of the fovea) and upstream (by sending outputs to several areas thought to be involved in social cognition; Powell et al. 2018). It is also natural to do the inverse move, and appeal to different functions to explain why different neural structures are indeed different: the fact that the striate cortex plays an important role in vision, for example, offers an immediate and satisfactory explanation for its retinotopical organization (cf. Tootell et al. 1998).

Now, carrying out these two roles flawlessly requires having an ideal or complete cognitive neuroscience, which it's nowhere near us. And this is exactly why the third epistemic role of structure-function mappings is relevant; that is, a heuristic role in generating and testing hypotheses to progress and refine neuroscientific theories. Considered diachronically, the quest for structure-function mappings should propel scientists into refining their categories of cognitive functions and neural structures, aiming at increasing the explanatory and predictive power of structure-function mappings (Price & Friston 2005; Poldrack & Yarkoni 2016). As we mentioned in §1, early structure-function mappings were rather simple (and simplistic), owing them the charge of resembling modern phrenology (Uttal 2001). Yet, the charge of phrenology can be dropped insofar these mappings are not taken as definitive theory but rather working hypotheses to be evolved into more nuanced mappings via a series of iterative steps. For instance, when it turned out that, beside faces, the FFA of expert perceivers is also activated by holistic recognition of stimuli such as chess games or radiological images, some scholars proposed a redefinition of its function from "face recognition" to "expertise-based recognition" (Bilalić et al. 2011; 2016). In this way, testing a structure-function mapping forces neuroscientists to come up with a novel, better cognitive function, mapping in a more fine-grained way on neural structures (Bechtel 2002; Francken et al. 2022).

Summing up: structure-function mappings should play a predictive role, an explanatory role and a heuristic role in contemporary cognitive neuroscience. And these roles are central to the practice of cognitive neuroscience. Any theoretical framework preventing a structure-function mapping from satisfactorily playing these roles deprives cognitive neuroscientists of a valuable tool. We fear that this is the case with PP, which we shall now introduce.

3 - Predictive Processing

PP is a neurocomputational framework claiming that the brain's cognitive functioning can be understood in terms of a single computational objective, that of minimizing prediction error (see Friston 2005; Clark 2013, 2016; Hohwy 2013; Rao 2024). Thus, PP is - or at least strives to provide - a comprehensive, unificatory, and complete understanding of the brain and cognition in terms of a single computational operation (see Sprevak 2024). PP strives to be comprehensive, as it strives to offer an account of all cognitive - indeed all mental - phenomena, not just some of them (e.g. Hohwy 2015, 2020; Clark 2016, 2023). PP strives to be unifying, as it explains all facets of mentality as the product of a single algorithm, repeatedly executed by numerous different brain structures. And it strives to be complete, in that it wants to offer an explanation of cognition covering all explanatory levels, from computation to physical implementation. As such, PP (as we use the term here) is different from the simple usage of predictive coding algorithms to explain only some aspects of mentality, such as vision (e.g. Rao & Ballard 1999), and also from "Bayesian" cognitive psychology (e.g. Kersten, Mamassian & Juille 2004; Griffiths and Zaslavsky 2023). The former are not comprehensive and unifying, and the latter are usually not complete, sitting only at the computational/algorithmic level. Hence, these proposals are significantly different from the kind of proposals put forth under the banner of PP.

PP is canonically introduced considering perception and the hierarchical predictive coding algorithm. We will conform to this custom.[1] Thus, consider perception. Since sensory stimuli are noisy and informationally impoverished, they can't directly specify which worldly objects are generating them. Therefore, perceiving must amount to a form of (Bayesian) inference estimating the most likely cause of one's own sensory signals. PP suggests this estimate is drawn inverting a generative model thanks to a (hierarchical) predictive coding algorithm (Friston 2005; Clark 2013, 2016; Hohwy 2013).

Generative models are data structures capturing the probabilistic linkages between observations (in this case, sensory inputs) and their hidden causes (the objects and events generating them), which are also able to predict the observations each hidden cause likely produces. According to PP, such models are hierarchically structured, so as to capture the hierarchical causal structure of the external world producing our sensory inputs. Computationally, this just means that each hierarchical layer of the overall generative model has to treat the activity of the layer hierarchically below it as observations, trying to guess the inner causes of them (cf. Hinton 2007, Foster 2022).[2]

---

[1] We will follow the so-called "low road" to PP. There is also a "high road" to PP, based on certain transcendental conditions biological beings must satisfy to remain alive. See (Friston 2019) for a quick introduction to the "high road" and its comparison to the "low road". See also (Raja et al. 2021; Nave forthcoming) for a critical analysis.

[2] As it is customary, we will ignore the lateral (non hierarchical) complexity of generative models, about which see (Friston 2005).

PP claims that this hierarchy operates according to a (hierarchical) predictive coding algorithm (e.g. Rao & Ballard 1999).[3] Each hierarchical level (except the bottommost one directly observing the transducers) generates a signal, predicting the layer's observations (i.e. activity of the layer directly below it or incoming signals, in the case of the bottommost layer). Predictions are sent to the estimated layers, which confront them with their actual activity (or the incoming sensory signals). If a mismatch is detected, then the lower levels generate a prediction error signal, which is conveyed to the higher (predicting) layers, forcing them to revise their predictions so as to minimize the incoming prediction error. When error reaches a minimum throughout the hierarchy, then the most likely cause of the observation has been identified, in a way that inverts the linkages from observation to causes of the generative model.

Perception, however, is only one of the cognitive processes realized in this way. Since PP aims at being comprehensive, it claims that all mental phenomena are the result of prediction error minimization (e.g., Friston 2009, 2010; Hohwy 2015; Clark 2016, 2023; Spratling 2016; Rao 2024).[4] Action, for example, minimizes prediction error by bringing about the predicted sensory singlas thanks to movement (cf. Adams et al. 2013a). Attention minimizes the weight (i.e. impact on processing) of prediction errors deemed noisy and unreliable - thereby allowing only informative reliable prediction error spread in the cerebral hierarchy, minimizing the overall quantity of error signal (cf. Feldman & Friston 2010). Learning consists in the minimization of prediction error through adjustments of the parameters of the generative model, so as to make it yield less prediction error in the future (Friston 2005, 2010). Emotion consists in the prediction and control of inner interoceptive predictions, again aimed at prediction error minimization (Seth & Friston 2016, Barrett 2017). Social cognition consists in the minimization of prediction errors concerning someone else's behavior and behavioral causes (Friston & Frith 2015a, b). The list may continue (e.g. Spratling 2016; Hohwy 2015; 2020), but the main message of PP should now be clear: all mental phenomena boil down to prediction error minimization, one way or another.

The above provided a very schematic presentation of PP at the level of the algorithm. Recall, however, that PP aims to be complete, and speak also of the implementation level (Sprevak 2024). Hence defenders of PP project - more or less directly - this algorithm directly onto the brain, and especially the cortex (Mumford 1992; Friston 2005, 2009, 2010; Shipp 2016). For, the entire brain is supposed to be (cf. Friston 2013) - or, more properly, to realize - the generative model. Each hierarchically ordered cortical area, from "lower" primary sensory and motor areas to "higher" multimodal areas, realizes a layer of the hierarchical model. Top-down, descending cortico-cortical connections carry predictions, whereas bottom-up, ascending cortico-cortical

---

[3] The details of which differentiate various PP proposals, see (Sprevak unpublished a; Spratling 2017).

[4] Sometimes, this claim is expressed in terms of free-energy minimization rather than prediction error minimization. As it is customary in the literature, we will talk about free-energy models in terms of prediction error (e.g. Hohwy 2020; Clark 2016, 2023). On the differences between free-energy and prediction error minimization, see (Hohwy 2021). For an up to date, general introduction to the free-energy principle, see (Ramstead et al. 2023).

connections carry prediction errors (e.g. Markov et al 2014).[5] Superficial cortical layers (1-3) and the inner/deep granular layer (4) deal with precision and the messaging of prediction errors, whereas layer (5) generates descending predictions (cf. Mumford 1992; Shipp 2016). This, in extreme succinct terms, is the way in which the cortex implements the algorithmic scheme suggested by PP. But subcortical structures participate to PP too, often managing the weight of prediction errors (cf. Barrett & Simmons 2015; Kanai et al. 2015). Indeed, even the "computationally modest" retinal neurons are in the task of prediction inputs and signaling prediction errors (cf. Hosoya et al. 2005; Clark 2013). Similarly, the spinal cord has been depicted as minimizing prediction errors via the triggering of the reflex arcs needed to bring about predicted inputs (Adams et al. 2013a).

The moral, then, is that, on the view PP offers, each and every neural structure - be it a cortical network, an area, a cortical column (Bastos et al. 2012; Shipps 2016) or even and individual neuron (Isomura & Friston 2018; Isomura et al. 2023) - is in the task of minimizing prediction error. And so, PP appears to entail a all-to-one structure-function mapping: all neural structures perform a single function via a single algorithm; they all minimize prediction errors (cf. Friston 2009, 2010; Hohwy 2015; Clark 2023).

This all-to-one mapping is also central to PP's aims of comprehensiveness, unity and completeness. Insofar PP wants to be unifying, it must claim that a same algorithmic procedure is implied; and insofar it wants to be comprehensive, it can't but claim that all cognitive processes are the product of these algorithmic procedures. And given that PP wants to be complete - and sink its explanatory teeth also down to the implementation level, it can't but assign all different networks and neural structures only one function; namely that of minimizing the prediction error they receive.

4 - All-to-one mappings and equipotential brains

There is no doubt that the PP story canvassed above captures *something* about brain structure and functioning. It captures, for example, the well known (and independently established) fact that neuronal responses are always "contextualized" and partially driven by the overall brain activity, and are not "sheer" responses to the presentation of stimuli (e.g. Mesulam 1998). The multilayered nature of the generative model also captures the hierarchical structure of the cortex; and the independence of variables tracked in the generative model may also be reflected in the gross anatomical structure of the brain. The distinction between "what" and "where" streams, for example, may be a result of the fact that variables for location and identity of objects rarely influence each other (cf. Friston 2013).

All these explanations, however, concern some "global" aspect of brain structure, rather than the features of any individual neural structure, or how a single cognitive function can be mapped onto its neurobiological basis. So, they do not concern structure

---

[5] Through some PP models escape this scheme (e.g. Heeger 2017), and even in these who do not escape it, the interpretation of ascending and descending connections is more contested than what any standard interpretation of PP typically acknowledges (Orlandi & Lee 2019; Cao 2020; Walsh et al. 2020).

function mappings as typically understood in the cognitive brain sciences. How does PP fare in this regard?

In some cases, it performs well. Consider, to start, the *explanatory* role structure-function mappings are supposed to play. Neural structures should explain how functions are carried out, and the need to perform these functions explain why we observe these specific structures. Prediction error minimization does, at times, play this role. For example, ascribing the function of minimizing prediction error explains why the motor cortex has an agranular architecture.[6] The fourth, "granular" layer of the cortex is said to encode prediction error. But the motor cortex *must not encode it*. For, encoding it would prevent movement, as it would force the motor cortex to change predictions to fit the data, rather than to minimize prediction error via movement (cf. Adams *et al.* 2013, Shipp *et al.* 2013). In the case at hand, then, the ascribed function explains why we observe this particular neural structure, and the observed neural structure explains how the function is carried out. A similar line of thought can further explain the agranular nature of the limbic region, which controls (*via* prediction error minimization) our interoceptive stream of input (Barrett & Simmons 2017). Again, their role is that of keeping that stream *in check*, within *predicted* bounds of viability, which must not be altered, on the pain of developing serious and chronic conditions. Hence the need to minimize prediction error changing the input, rather than the predictions.

Similarly, PP can straightforwardly map attention to the brain, pointing to the series of mechanisms to selectively boost or diminish the impact of ascending neural signals (i.e. prediction errors, see Friston 2012). More generally, as noticed by Sprevak (unpublished b), PP often tends to appeal to facts about the physical implementation of the prediction error algorithm to account for a variety of *exceptions* from the standard (perception-like) prediction error minimization story and its obvious implications. In all these cases, the all-to-one mapping entailed by PP satisfies the explanatory role structure-function mappings are supposed to play.

These explanations are real and important. Yet, they do not concern "standard" cases of prediction error minimization in which predictions are altered so as to fit the incoming input. Indeed, the all-to-one structure function mapping *fails* to play the relevant predictive, explanatory and heuristic role in such standard cases. Hence, the successes of the all-to-one structure function mapping in playing these roles are exceptions, lying on a fairly large pile of failures.

Consider, first, the *predictive* role structure-function mappings are generally supposed to play. Knowing that a structure plays a given function should enable us to predict its (differential) activation in given contexts or tasks. It should also allow us to revert the inferential arrow, to infer what context or task most likely generated any activation pattern.

However, according to PP *each* structure plays the *same* role any other structure plays. So, it offers us no reason to predict that visual tasks should mainly activate visual cortices, nor that activations of visual cortices will be strongly suggestive of the

---

[6] Though note that the agranularity of the motor cortex is not uncontroversial. See for example (Yamawaki et al. 2014).

execution of visual tasks. Indeed, PP offers us no reason to expect *differential* activations at all. If really all structures perform the same function, we should expect them to be *equally active* in every task. Thus, we shouldn't expect the *task-specific* patterns of activation we factually observe.

PP does not fare better when it comes to the *explanatory* role structure-function mappings are supposed to play. It is true that ascribing the function of minimizing prediction error allows us to explain the inner organization of certain neural structures in a more or less direct manner (as seen above with agranualr cortices). And it is also true that PP may allow us to seamlessly explain the features of structures that are repeated throughout the brain, such as cortical columns (Bastos *et al*. 2012; Shipp 2016). Yet, given that PP ascribes the *same* function to all neural areas, it's very difficult to see how this function could explain the *variety* of neural structures we actually observe. Broadman's brain parcellation for *Homo sapiens* listed more than 40 anatomically differentiated areas (Broadman 1909/2006), and modern counts number around 180 anatomically differentiated areas per hemisphere (Glasser *et al.* 2016). This sort of variety surely doesn't seem a negligible explanandum (cf. Rathkopf 2013, p. 12), yet it is hard to see how PP can account for it. Why do we have *so many* different neural structures, if they all perform the same role? And how can all these different structures ultimately perform the same function? Relatedly, PP seems to make it mysterious why neural activations are selective and task specific. Why, if all neural structures perform the exact same function, are they selectively activated by specific tasks? As far as we can see, PP leaves these questions unanswered. Hence the structure-function mapping it entails seems unable to play the desired explanatory role.

PP does not fare better when it comes to the *heuristic* role. PP is often regarded as a research programme which brought beneficial epistemic effect on the mind sciences as a whole (see Sprevak 2024). It provided a heuristic framework weaving together various distinct threads of research in a coherent and attractive picture (e.g. Clark 2016; Allen & Friston 2018) and inspired various computational models in robotics, artificial intelligence and the study of complex systems (cf. Tani 2016; Parr *et al*. 2022; Ramstead *et al*. 2023). However, such beneficial epistemic effects have not yet yielded many fruits in terms of working hypotheses of structure-function mappings. Recall (**§2**): to play the relevant heuristic role, structure-function mappings should lead to the progressive refinement of our functional and structural labels. But PP *cannot* refine the functional labels it provides, as long as it remains in the grip of its comprehensive, unificatory goals. And whilst PP is compatible with a variety of taxonomies of neural structures, these taxonomies cannot interact in any interesting way with the functional labels PP provides. For, no matter how neural structures are taxonomized, PP will project onto each neural structure the same function of minimizing prediction error. Thus, it is again hard to see how the all-to-one structure-function mapping PP entails may play the desired heuristic role. To be comprehensive, unificatory and complete, PP can't provide any functional tag other than prediction error minimization. But this single functional tag interacts minimally with structural taxonomies, in a way that seems to prevent the former from playing any heuristic role when it comes to discovering the latter. Nor new functional taxonomies can refine PP, at least not without compromising its aspirations of comprehensiveness, unification and completeness (see §6).

So, the all-to-one structure-function mapping entailed by PP is not well suited to play the heuristic roles structure-function mappings are supposed to play. Worse still, that mapping paints a very problematic picture of the brain as a whole, as a massively degenerate, and indeed equipotential, organ.

Degeneracy is the property of systems whereby a single function is played by structurally different components (cf. Edelman & Gally 2001). The immune system is a prime example of degeneracy: the elimination of virus and bacteria can be subserved by various types of structurally distinct cells like phagocytes, lymphocytes, mastocytes and other types of white blood cells.

Some degeneracy makes good biological sense, for it allows a system to be robust. We survive most pathogens because the immune system has various different "defense mechanisms" which may compensate for the deficiencies of each other. A pathogen unscathed by phagocytes may be eradicated by lymphocytes, and vice-versa.

Yet, in the picture PP proposes, the brain is not just degenerate - it is *entirely and massively* degenerate: *all* structurally different components do *exactly* the same thing. And this makes *less* biological sense.

A first problem concerns the *metabolic cost* of our brains. Whilst the PP algorithm is metabolically efficient (Ali *et al*. 2022; Hechler *et al* 2023), human brains are *still* metabolically pricey, accounting for about 20% of bodily energy consumption (Raichle 2006). Having a brain like ours, then, is a significant metabolic investment, which requires some sort of justification (see Sterling & Laughlin 2015). Such a justification can be easily provided by theories of brain functioning allowing different neural structures to play *different* functional roles: humans need to spend about 20% of their metabolic budget in their big brain because a big brain can be anatomically differentiated, which makes them able to play all the *different* cognitive functions humans need to execute to thrive. PP, however, cannot provide this answer, as it does not allow for *different* cognitive functions to be assigned to neural areas. So, how can it justify the metabolic investment in big brains? As far as we can see, this question is left unanswered in the PP literature.

Massive degeneracy also makes it mysterious why our brains are constituted by *anatomically different* structures. Anatomical differentiation is often an indicator of functional differentiation: organs whose cells all play the same functional roles, such as the liver, tend not to be internally partitioned in different anatomical structures (McCaffrey 2015). According to PP, however, our neural structures are not functionally differentiated: they only play exactly the same functional role. So, PP would lead us to expect the brain to be a *homogeneous* organ with no appreciable internal structure - which it clearly isn't.

Relatedly: why if the anatomically different structures composing the brain have the same function, they *do not* equally participate in all cognitive functions? Why do we see diversified, task- and function- specific patterns of activation in the brain? Why, if the V1 and the supplementary motor area (SMA) play exactly the same functional role, does it never happen that V1 is selectively recruited in motor tasks and the SMA is recruited to perform visual tasks? These questions are easily answered if we assume that V1 and SMA play different functional roles. But how can this question be answered, if we assume that V1 and SMA have the same function? Sadly, as far as we can see, PP leaves us with no guidance on this matter.

This last point, we think, should be urgently addressed by defenders of PP. For, as things stand, it is hard not to conclude that, from the point of view PP offers, the brain is an *equipotential* organ. Not only PP does not predict the task-specific differential patterns of activation we *in fact* observe, it also predicts *unspecific* activations of *all* neural structures in *all* tasks. And this is precisely the core prediction of *equipotentialism* (Flourens & Nadel 1824).

But *equipotentialism is simply wrong*. Sure, the brain is *plastic*: the functional role of many neural structures is not set in stone, and many neural structures often reorganize themselves either to perform their function better or to acquire some novel function.[7] But plasticity does not offer any foothold to either all-to-one mappings nor to equipotentialism. Plasticity allows neural structures to play *diverse* functions, opposing the all-to-one structure-function mapping. And neural plasticity is *constrained* by several structural and anatomical features. It is not the case that *any* neural structure can assume *any* function with *any* arbitrary degree of success. So, for example, whilst rewiring the visual cortex to the auditory thalamus in newborn ferrets allows the visual cortex to reorganize as a make-shift auditory cortex and forces the thalamus to take up the role of the visual cortex (Sharma *et al*. 2000), the thalamus *cannot* completely substitute for the visual cortex, as it is, for example, unable to perform the discriminations needed to determine the orientation of stimuli (von Melchener *et al*. 2000).

The all-to-one structure-function mapping suggested by PP thus appears to fail to play the desired epistemic roles structure-function mappings are supposed to play (**§2**). Worse still, it suggests an equipotential picture of the brain, a picture that we know is factually wrong.[8]

This seems like a hard blow to PP. A fatal blow, perhaps? Let us consider some of the ways in which defenders of PP might react to our claim.

5 - Is the threat to PP real?

Some defenders of PP may wish to resist our analysis. Let us consider some possible lines of resistance, and show that they do not really shield PP from our worries.

---

[7] See, however, (Makin & Krakauer 2023) for a powerful argument to the contrary.

[8] These problems are even thornier for embodied readings of PP (e.g. Clark 2016, Kersten forthcoming) and views of PP closer to the expansive framework of the free-energy principle (e.g. Friston & Stephan 2007). Embodied readings of PP argue that, at least at times, non-neural bodily structures play the role of minimizing prediction error. When it comes to the free-energy principle, one core claim is that everything an organism does (from cognizing to eating) minimizes free-energy, which is (under certain assumptions) an upper bound of prediction error. These readings make the link between structure and function even more tenuous, and the issue of degeneracy even more pressing. Why, if non-neural structures have the function of minimizing prediction error/free-energy too, do we have brains at all? Wouldn't non-neuronal tissue be enough? And how do non-cognitive functions help us understand the structure of our non-neural organs, if these organs too are just in the task of minimizing free-energy?

A first objection contends that we have attacked a strawman because our reconstruction of PP is misguided. PP simply does not really assign the same function to all neural structures - and so the problems we've diagnosed are simply not real.

Sure, some applications of predictive coding are limited to only some neural functions or structures (e.g. Rao & Ballard 1999). But we are not talking about those. We are talking about PP, which strives for comprehensiveness, unity and completeness. For example Clark wrote:

> Matter, when organized so that it cannot help but try (and try, and try again) to successfully predict the complex plays of energies that are washing across its energy-sensitive surfaces, has many interesting properties. Matter, thus organized, turns out, as we'll see, to be ideally positioned to perceive, to understand, to dream, to imagine, and (most importantly of all) to act. Perceiving, imagining, understanding, and acting are now bundled together, emerging *as different aspects and manifestations of the same underlying prediction-driven, uncertainty-sensitive, machinery*. (Clark 2016, p xiv, emphasis added)

It is natural to read Clark as claiming that many (if not all, as the book seems to suggest) cognitive functions boil down just to prediction error minimization, and so that it is sufficient for the brain to minimize prediction error to give rise to our mental life. That is, Clark is claiming that many (maybe all) different neural structures need only to perform that single function; he's suggesting an all-to-one mapping.

Similarly Spratling (2016) relies on PP to model cognition in general, Friston (2009; 2010) offers it as an "unified theory" of brain and cognition, and Hohwy (2015) explicitly states that PP is set off to explain the mind in its globality. It's hard to see how these proclaims could coherently be read without also endorsing the view that all the brain does is to minimize prediction error, and the consequent all-to-one mapping.

Moreover, defenders of PP appear keen to downplay or even neglect functional differences in cortical areas:

> The primary motor cortex is no more or less a motor cortical area than striate (visual) cortex. The only difference between the motor cortex and visual cortex is that one predicts retinotopic input while the other predicts proprioceptive inputs from the motor plant (Friston, Mattout and Kilner 2011: 138).

A natural (literal) and charitable interpretation of this passage is that it denies any functional difference between motor and visual areas. Indeed, the passage suggests that they both have the function of predicting certain inputs and minimizing the error relative to these predictions. The same function has also been assigned to structures in the "emotional brain" such as the insula and the amygdala (cf. Seth 2015; Seth & Friston 2016; Barrett & Simmons 2017). The only difference between all these various neural structures seems only the sensory origin of the prediction error they minimize. So, we are not attacking a strawman here.

A second objection claims that we have been using a functional taxonomy that is too coarse-grained. PP does not claim that neural structures have the function of minimizing prediction error in general. Rather, it claims that neural structures have the function of minimizing specific streams of prediction error. For example, the visual cortices have the function of minimizing visual prediction error, whereas auditory ones minimize auditory prediction error, and the motor and somatosensory cortices minimize somatosensory prediction errors, and so forth. This allows PP to have a form of functional differentiation (avoiding the problems connected with the all-to-one mapping) while maintaining comprehensiveness and unity unaltered.

In reply, we note two flaws of this move. One is that the move still does not offer enough functional differentiation. V1, V2, the FFA would all be assigned the same function of minimizing visual prediction error. But how can that function explain their anatomical differences, or predict the specific task-dependent patterns of activation we observe? Why is the FFA rather than V1 differentially activated by faces? Why are V1 and the FFA anatomically distinct, if they perform the same function? We are back to confronting the problems we highlighted above, only at a smaller, modality-specific scale. Moreover, what sort of function would be assigned to non-sensory, associative cortices such as the prefrontal cortex? Such cortices do not seem to minimize any specific stream of sensory prediction error. The second flaw is instead that even primary sensory motor areas do not appear to deal with signals coming only from one modality (see for example Bell et al. 2019; Heimler & Amedi 2020; Rosemblun et al 2016; Murray et al 2016; Ghazanfar & Schroeder 2006). But if so, then it seems we have no good reason to claim, say, that the visual cortices minimize visual prediction errors. We are thus left only with the original, “coarse-grained” all-to-one structure function mapping.

A related objection contends that PP assigns a single function to all brain areas only at a very high, and unhelpful, level of abstraction. At an appropriate level of abstraction PP would assign different functions to different neural structures. Compare: in very general terms, we can say that all cells in our body have the function of maximizing our fitness. This does not mean, of course, that that is their only function, nor the most relevant one in understanding what different types of cells do and how they do keep us alive.

The analogy, however, doesn't really add up: surely, we can say that all our cells do is maximizing our fitness, and we agree that this is far from illuminating. And in fact, as far as we know, no biologist is proposing that as a framework to understand cell function and structure. Defenders of PP, however, do claim that observing the brain and brain structures through the lenses of prediction-error minimization is genuinely illuminating (see, for example, Friston 2009, 2010; Howhy 2013, 2015; Clark 2013, 2016, 2023). So reading their claims as pitched at such a high level of abstraction seems to trivialize them.

At this point, the objector would likely argue that - at least in the free-energy variants of PP - the prediction error minimization has been unpacked in a variety of different computational operation, such as, for example, complexity/accuracy trade offs in the internal model servicing learning, risk/uncertainty trade offs servicing exploratory behaviors, and the minimization of a new, related quantity called expected free-energy (see Friston et al. 2023 for a global view of these changes). So, it is no longer true that the only relevant function is prediction error (or free-energy) minimization.

These developments are, for us, a step in the right direction, and do something to ameliorate the problems we have canvassed above. In particular, they are in and by themselves sufficient to avoid the all-to-one structure-function mapping, and thus the charge of equipotentialism. Still, PP would be left with a handful of functions to be paired to about 180 cortical areas and subcortical structures. There is little hope that these few functions will allow us to appropriately explain and highlight the diversity of neural structures. And in any case, in order to be psychologically meaningful, these functional labels will still have to be reconnected with a behaviorally transparent redescription of phenomena (Francken & Slors 2017; Klein 2012).

Another objection attacks our (general) commitment to localizationism. Perhaps we shouldn't try to assign functions to neural structures at all. Perhaps we should do something different. We might, for example, try to assign each structure a "working" (that is, an intrinsic computational profile) and then investigate how these "workings" enable the structure to partake, together with other structures, in the execution of various cognitive processes (Bergeron 2007; Anderson 2010). Or we could, to give another example, assign each structure a different "propensity" or "personality" making it more or less apt to cooperate with other structures in the execution of cognitive tasks (Anderson 2014).

Now, it could seem that abandoning localizationism offers a straightforward solution to our problem. If we altogether abandon the idea of structure-function mappings, we also, eo ipso, abandon the problems the all-to-one mappings generate. Yet the escape from these problems is, alas, only illusory. Non-localizationist proposals still need to call upon different "workings" or "propensities" to account for the anatomical differences in brain structures, and to explain why neural structures are selectively recruited in various tasks. Yet, it is far from clear whether predictive processing can provide those. If "workings" are the intrinsic computational profile of neural structures, then it's hard to see how PP can avoid to claim that all structures have the same "working", as they all minimize prediction error according to the same algorithm. Maybe PP could assign differently, empirically determined, "propensities" to various neural structures (based, for example, on their response profile). But this move would come at a high price, as it hinders the completeness and comprehensiveness. Completeness would be under threat because PP would need an account of how algorithmically identical prediction error minimization units end up having different "propensities"; that is PP would need a novel way to relate the algorithmic level of prediction error minimization to the implementation level of "propensities". Comprehensiveness would be under threat because a comprehensive neurocognitive account would have to mention the relevant "propensities", that are not part PP.

In sum, it seems that, in order to escape our argument, defenders of PP have to withdraw from its bolder ambitions of comprehensiveness, unity and/or completeness. Is that a price worth paying? As we will argue in the next and final section, probably so.

6 - Conclusions: prediction without comprehensiveness, unity and completeness

We have argued that the structure-function mapping entailed by PP is unable to play the epistemic roles structure-function mappings are supposed to play. Indeed, PP paints a puzzling picture of the brain as a massively degenerate, equipotential organ - a picture we have excellent reasons to believe is false.

It's important to notice, however, that the problems we have been diagnosing do *not* strictly speaking come from the theoretical constructs PP deploys to explain the brain and cognition. These problems are *not* rooted in prediction error minimization *per se*. Rather, they stem from PP's aspiration to comprehensive, unity and completeness. It is because PP has these aspirations that it is forced to ascribe *all* neural areas the *same* function, and thus in putting forth its all-to-one mapping.

This means that, from the perspective of PP, our conclusions really are a mixed bag. On the one hand we have repeatedly noticed that PP *can do* - and in fact often does - some relevant explanatory work. It can, for example, explain why certain cortices boast an agranular structure (Shipp et al. 2013). PP seems also well positioned to explain recurrent computational *motifs* throughout the brain, such as the presence of cortical columns (Bastos *et al*. 2012; Shipp 2016). These explanations are real and important, and testify nicely that the explanatory tools of PP can prove useful in the mind sciences.

However, we wish to insist that these tools alone *do not* provide a global, unified explanation of the brain and cognitive processing. Indeed, as we've argued, when used alone and applied indiscriminately, the explanatory tools of PP yield a picture of the brain as an equipotential organ - a picture we know is false.

The natural moral to draw, then, is that PP, useful as it may be, cannot be the grand unified theory of the brain and cognition it aspires to be: its goals of comprehensiveness, unity and completeness are beyond its own epistemic reach.

This isn't, in and by itself, a novel conclusion. Following independent lines of arguments, other scholars (e.g. Colombo & Wright 2017; Litwin & Miłkowski 2020) have also highlighted that PP falls short of being the comprehensive, unificatory theory it promises to be. Our work diagnoses the same problem from a new, previously unexplored, angle. That several researchers reached the same conclusion walking different paths suggest that this conclusion is robust: there are several, theoretically independent, reasons to think that PP can't be the complete, comprehensive and unificatory theory it is typically publicized to be. Suppose, now, that PP renounces its ambitions as a comprehensive, unificatory and complete theory of cognition. What would follow for structure-function mappings? We think that the answer is, minimally, a form of *functional pluralism* (McCaffrey & Wright 2022). We can still continue to assign each and every neural structure the function of minimizing prediction error (supposing, for the sake of discussion, that all neural structures *do* minimize prediction error); but, in addition to that, we must also assign differentiated, specialized functions to individual neural structures which clarifies how they contribute to behavior. Notice that, in principle, there is nothing puzzling about this sort of assignment of functions. Each component of an internal combustion engine has the function of *making the car go* (see Klein 2012). But each component also has some individual function that is uniquely played by that component (e.g. only the igniter has the function of igniting the fuel).

Such a pluralistic assignment of functions could allow us to dodge the problems we have highlighted above. For, the individualized functions (or functional propensities, if you prefer) assigned to each neural structure are *in and by themselves* sufficient to avoid casting the brain as an equipotential organ, to avoid massive degeneracy, and to play the epistemic roles of structure-functions mapping. Equipotentiality and massive degeneracy are avoided because neural structures are assigned (also) different functions. And the mapping from individual neural structures to specific functions does play the epistemic roles structure-function mappings are supposed to play as well as *any* other structure function mapping, given that, PP aside, all structure-function mappings map individual neural structures on different functions (even if the mapping is often many-to-many, see McCaffrey 2023, Westlin et al. 2023).

It is worth noticing, in closing, that PP itself is developing in a way that requires novel, different functional assignments. For example, to explain exploratory behaviors and curiosity within a PP framework, Kiverstein *et al* (2019) are forced to call upon a mechanism tracking the rate of prediction error minimization. Such a mechanism has not yet been mapped on any neural structure, but its mere presence within the theoretical apparatus of PP already introduces the need to assign functions *other* than prediction error minimization to at least some neural structures. Similarly, as noticed above, free-energy variants of PP have introduced a second quantity to minimize (so-called *expected* free-energy) and thus a second function that needs to be mapped on neural tissue. An analogous need for multiple functional assignments is emerging also from less theory-driven considerations. For example, Hull (2020) has recently reviewed the functionality of the cerebellum from a predictive processing perspective. Whilst the review takes the cerebellum as a predictor of sorts, the functions (*plural*) that are invoked in the review go well beyond *just* prediction error minimization. For example, Hull argues that the cerebellum minimizes *reward* prediction error, which is distinct from the *sensory* prediction error minimized in the cortex.[9] Similarly, whilst the cerebellum is cast as a forward (generative) model busy predicting the sensory consequences of our movements, cerebellar predictions are described as based on *motor commands* computed by the motor cortex, which is thus assigned a function *other* than prediction error minimization.[10] The cerebellum is also said to learn according to a non-trivial range of learning rules, which, again, go beyond the account of learning based *exclusively* on prediction error minimization that PP offers.

Thus it seems that our claim that PP (in its current form) is not a comprehensive, unifying and complete theory of the brain and cognition, and our consequent call for a form of functional pluralism, is less inimical to the *actual development* of predictive processing than the "official" PP rhetoric may suggest.

References

Adams, R. A., Shipp, S., & Friston, K. J. (2013). Predictions not commands: active inference in the motor system. Brain Structure and Function, 218, 611-643.

---

[9] On their difference, see again (Sprevak 2024).

[10] And indeed, it is worth noticing that versions of predictive processing insisting on unity, comprehensiveness and completeness try to eliminate both rewards (Friston et al 2012) and motor commands (Adams et al. 2013).

Ali, A., Ahmad, N., de Groot, E., van Gerven, M. A. J., & Kietzmann, T. C. (2022). Predictive coding is a consequence of energy efficiency in recurrent neural networks. Patterns, 3(12).

Allen, M., & Friston, K. J. (2018). From cognitivism to autopoiesis: towards a computational framework for the embodied mind. Synthese, 195(6), 2459-2482.

Anderson, M. L. (2010). Neural reuse: A fundamental organizational principle of the brain. Behavioral and brain sciences, 33(4), 245-266.

Anderson, M. L. (2014). After phrenology: Neural reuse and the interactive brain. The MIT Press.

Anderson, M. L., Kinnison, J., & Pessoa, L. (2013). Describing functional diversity of brain regions and brain networks. Neuroimage, 73, 50-58.

Barrett, L. F. (2017). The theory of constructed emotion: an active inference account of interoception and categorization. Social cognitive and affective neuroscience, 12(1), 1-23.

Barrett, L. F., & Simmons, W. K. (2015). Interoceptive predictions in the brain. Nature reviews neuroscience, 16(7), 419-429.

Bastos, A. M., Usrey, W. M., Adams, R. A., Mangun, G. R., Fries, P., & Friston, K. J. (2012). Canonical microcircuits for predictive coding. Neuron, 76(4), 695-711.

Bechtel, W. (2002). Decomposing the mind-brain: A long-term pursuit. Brain and Mind, 3, 229-242.

Bechtel, W. (2008). Mental mechanisms: Philosophical perspectives on cognitive neuroscience. Taylor & Francis.

Bergeron, V. (2007). Anatomical and functional modularity in cognitive science: Shifting the focus. Philosophical Psychology, 20(2), 175-195.

Bell, L., Wagels, L., Neuschaefer-Rube, C., Fels, J., Gur, R. E., & Konrad, K. (2019). The cross-modal effects of sensory deprivation on spatial and temporal processes in vision and audition: A systematic review on behavioral and neuroimaging research since 2000. Neural Plasticity, 2019, 9603469.

Bilalić, M., Langner, R., Ulrich, R., & Grodd, W. (2011). Many faces of expertise: fusiform face area in chess experts and novices. Journal of neuroscience, 31(28), 10206-10214.

Bilalić, M., Grottenthaler, T., Nägele, T., & Lindig, T. (2016). The faces in radiological images: fusiform face area supports radiological expertise. Cerebral Cortex, 26(3), 1004-1014.

Broadmann, K. (1909/2006). Vergleichende Lokalisationslehre der Grosshirnrinde in ihren Prinzipien dargestellt auf Grund des Zellenbaues, Johann Ambrosius Barth Verlag, Lipsia. English translation: Localization in the cerebral cortex. L. J. Garvey Trans, New York, NY: Springer.

Burnston, D. C. (2016). A contextualist approach to functional localization in the brain. Biology & Philosophy, 31, 527-550.

Cao, R. (2020). New labels for old ideas: Predictive processing and the interpretation of neural signals. Review of Philosophy and Psychology, 11(3), 517-546.

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. Behavioral and brain sciences, 36(3), 181-204.

Clark, A. (2016). Surfing Uncertainty. Oxford University Press.

Clark, A. (2023). The experience machine. Penguin

Colombo, M., & Wright, C. (2017). Explanatory pluralism: An unrewarding prediction error for free energy theorists. Brain and Cognition, 112, 3-12.

Craver, C. F. (2007). Explaining the brain: Mechanisms and the mosaic unity of neuroscience. Clarendon Press.

De Brigard, F., & Gessell, B. (2024). The mirage of big-data phrenology. https://doi.org/10.1086/732152

Edelman, G. M., & Gally, J. A. (2001). Degeneracy and complexity in biological systems. Proceedings of the National Academy of Sciences, 98(24), 13763–13768. https://doi.org/10.1073/pnas.231499798

Feldman, H., & Friston, K. J. (2010). Attention, uncertainty, and free-energy. Frontiers in human neuroscience, 4, 215.

Flourens, M. J. P., & Nadel, L. (1824). Recherches experimentales sur les propietes et fonctions due systeme nerveus dans les animaus vertebres. Paris. Polster, MR, Nadel, L, and Schacter, DL (1991) Cognitive neuroscience analysis of memory: A historical perspective. Journal of Cognitive Neuroscience, 3(2), 95–116. https://doi.org/10.1162/jocn.1991.3.2.95

Foster, D. (2022). Generative deep learning. O'Reilly Media, Inc.

Francken, J. C., & Slors, M. (2018). Neuroscience and everyday life: Facing the translation problem. Brain and Cognition, 120, 67-74.

Francken, J. C., Slors, M., & Craver, C. F. (2022). Cognitive ontology and the search for neural mechanisms: three foundational problems. Synthese, 200(5), 378.

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

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

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

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

Friston, K. (2013). Active inference and free energy. Behavioral and brain sciences, 36(3), 212.

Friston, K. (2019). Beyond the desert landscape. In Andy Clark & His Critics (ed. M. Colombo, M. Stapleton & L. Irvine). Oxford University Press.

Friston, K., Adams, R., & Montague, R. (2012). What is value—accumulated reward or evidence?. Frontiers in neurorobotics, 6, 11.

Friston, K., Da Costa, L., Sajid, N., Heins, C., Ueltzhöffer, K., Pavliotis, G. A., & Parr, T. (2023). The free energy principle made simpler but not too simple. Physics Reports, 1024, 1-29.

Friston, K. J., & Frith, C. (2015a). A duet for one. Consciousness and Cognition, 36, 390–405.

Friston, K. J., & Frith, C. (2015b). Active inference, communication and hermeneutics. Cortex, 68, 129–163

Friston, K. J., & Stephan, K. E. (2007). Free-energy and the brain. Synthese, 159, 417-458.

Ghazanfar, A., & Schroeder, C. (2006). Is neocortex essentially multisensory? Trends in Cognitive Sciences, 10(6), 278–285

Glasser, M. F., Coalson, T. S., Robinson, E. C., Hacker, C. D., Harwell, J., Yacoub, E., ... & Van Essen, D. C. (2016). A multi-modal parcellation of human cerebral cortex. Nature, 536(7615), 171-178.

Griffiths, T. L., & Zaslavsky, N. (2023). Bayesian approaches to color category learning. In Encyclopedia of Color Science and Technology (pp. 85-89). Cham: Springer International Publishing.

Haueis, P. (2018). Beyond cognitive myopia: a patchwork approach to the concept of neural function. Synthese, 195(12), 5373-5402.

Hechler, A., de Lange, F., & Riedl, V. (2023). The energy metabolic footprint of predictive processing in the human brain. bioRxiv, 2023-12.

Heeger, D.J. (2017). Theory of cortical function. PNAS 114(8): 1773-1782.

Heimler, B., & Amedi, A. (2020). Are critical periods reversible in the adult brain? Insights on cortical specialisations based on sensory deprivation studies. Neuroscience and Biobehavioral Reviews, 116, 494–507

Hinton, G. (2007). Learning multiple layers of representations. Trends in Cognitive Sciences, 11(10), 428–434

Hohwy, J. (2013). The Predictive mind. Oxford University press.

Hohwy, J. (2015). The neural organ explains the mind. In T. Metzinger & J. M. Windt (Eds). Open MIND: 19(T). Frankfurt am Main: MIND Group. doi: 10.15502/9783958570016

Hohwy, J. (2020). New directions in predictive processing. Mind & Language, 35(2), 209-223.

Hohwy, J. (2021). Self-supervision, normativity and the free energy principle. Synthese, 199(1-2), 29-53.

Hosoya, T., Baccus, S. A., & Meister, M. (2005). Dynamic predictive coding by the retina. Nature, 436(7047), 71-77.

Hull, C. (2020). Prediction signals in the cerebellum: beyond supervised motor learning. elife, 9, e54073.

Isomura, T., & Friston, K. (2018). In vitro neural networks minimise variational free energy. Scientific reports, 8(1), 16926.

Isomura, T., Kotani, K., Jimbo, Y., & Friston, K. J. (2023). Experimental validation of the free-energy principle with in vitro neural networks. Nature Communications, 14(1), 4547.

Kanai, R., Komura, Y., Shipp, S., & Friston, K. (2015). Cerebral hierarchies: predictive processing, precision and the pulvinar. Philosophical Transactions of the Royal Society B: Biological Sciences, 370(1668), 20140169.

Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: a module in human extrastriate cortex specialized for face perception. Journal of neuroscience, 17(11), 4302-4311.

Kersten, D., Mamassian, P., & Yuille, A. (2004). Object perception as Bayesian inference. Annu. Rev. Psychol., 55(1), 271-304.

Kersten, L. (forthcoming). Predicting the Body or Embodied Prediction? New Directions in Embodied Predictive Processing In Larry Shapiro & Shannon Spaulding (eds.), Routledge Handbook of Embodied Cognition (2nd Edition). New York: Routledge. pp. 1-17

Kiverstein, J., Miller, M., & Rietveld, E. (2019). The feeling of grip: novelty, error dynamics, and the predictive brain. Synthese, 196, 2847-2869.

Klein, C. (2012). Cognitive ontology and region-versus network-oriented analyses. Philosophy of Science, 79(5), 952-960.

Litwin, P., & Miłkowski, M. (2020). Unification by fiat: arrested development of predictive processing. Cognitive Science, 44(7), e12867.

Makin, T. R., & Krakauer, J. W. (2023). Against cortical reorganisation. Elife, 12, e84716.

Markov, N. T., Vezoli, J., Chameau, P., Falchier, A., Quilodran, R., Huissoud, C., Lamy, C., Misery, P., Giroud, P., Ullman, S., Barone, P., Dehay, C., Knoblauch, K., & Kennedy, H. (2014). Anatomy of hierarchy: Feedforward and feedback pathways in macaque visual cortex. J. Comp. Neurol., 522(1): 225–259

McCaffrey, J. B. (2015). The brain’s heterogeneous functional landscape. Philosophy of Science, 82(5), 1010-1022.

McCaffrey, J. B. (2023). Evolving Concepts of Functional Localization. Philosophy Compass, e12914.

McCaffrey, J., & Wright, J. (2022). Neuroscience and Cognitive Ontology: a Case for Pluralism. In F. De Brigard & W. Sinnott-Armstrong (Eds.), Neuroscience and Philosophy. MIT Press. https://doi.org/10.7551/mitpress/12611.001.0001

Mesulam, M. M. (1998). From sensation to cognition. Brain: a journal of neurology, 121(6), 1013-1052.

Mumford, D. (1992). On the computational architecture of the neocortex: II The role of cortico-cortical loops. Biological cybernetics, 66(3), 241-251.

Murray, M. M., Thelen, A., Thut, G., Romei, V., Martuzzi, R., & Matusz, P. J. (2016). The multisensory function of the human primary visual cortex. Neuropsychologia, 83, 161–169.

Nave, K. (forthcoming). A drive to survive. The MIT Press.

Orlandi, N., & Lee, G. (2019). How radical is predictive processing? In Andy Clark & His Critics (ed. M. Colombo, M. Stapleton & L. Irvine). Oxford University Press.

Parr, T., Pezzulo, G., & Friston, K. J. (2022). Active inference: the free energy principle in mind, brain, and behavior. MIT Press.

Poldrack, R. A. (2010). Mapping mental function to brain structure: how can cognitive neuroimaging succeed?. Perspectives on psychological science, 5(6), 753-761.

Poldrack, R. A., & Yarkoni, T. (2016). From brain maps to cognitive ontologies: informatics and the search for mental structure. Annual review of psychology, 67, 587-612.

Poth, N. (2022). Schema-centred unity and process-centred pluralism of the predictive mind. Minds and Machines, 32(3), 433-459.

Price, C. J., & Friston, K. J. (2005). Functional ontologies for cognition: The systematic definition of structure and function. Cognitive Neuropsychology, 22(3-4), 262-275.

Powell, L. J., Kosakowski, H. L., & Saxe, R. (2018). Social origins of cortical face areas. Trends in cognitive sciences, 22(9), 752-763.

Raichle, M. E. (2006). The brain's dark energy. Science, 314(5803), 1249-1250.

Raja, V., Valluri, D., Baggs, E., Chemero, A., & Anderson, M. L. (2021). The Markov blanket trick: On the scope of the free energy principle and active inference. Physics of Life Reviews, 39, 49-72.

Ramstead, M. J., Sakthivadivel, D. A., Heins, C., Koudahl, M., Millidge, B., Da Costa, L., ... & Friston, K. J. (2023). On Bayesian mechanics: a physics of and by beliefs. Interface Focus, 13(3), 20220029.

Rao, R. P. (2024). A sensory–motor theory of the neocortex. Nature Neuroscience, 27(7), 1221-1235.

Rao, R., & Ballard, D. (1999). Predictive coding in the visual cortex: A functional interpretation of some extra-classical receptive field effects. Nature Neuroscience, 2(1), 79–87.

Rathkopf, C. A. (2013). Localization and intrinsic function. Philosophy of Science, 80(1), 1-21.

Rosenblum, L. D., Dias, J. W., & Dorsi, J. (2016). The supramodal brain: Implications for auditory perception. Journal of Cognitive Psychology, 29(1), 65–87

Schalk, G., et al. (2017). Facephenes and rainbows: Causal evidence for functional and anatomical specificity of face and color processing in the human brain. Proceedings of the National Academy of Sciences, 114(46), 12285-12290.

Seth, A. K. (2015). The cybernetic bayesian brain. In T. Metzinger, J. Windt (eds.). Open MIND, 35. Frankfurt am Main, The MIND Group. https://doi.org/10.15502/9783958570108.

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

Sharma, J., Angelucci, A., & Sur, M. (2000). Induction of visual orientation modules in auditory cortex. Nature, 404(6780), 841-847.

Shipp, S. (2016). Neural elements for predictive coding. Frontiers in psychology, 7, 1792.

Shipp, S., Adams, R. A., & Friston, K. J. (2013). Reflections on agranular architecture: predictive coding in the motor cortex. Trends in neurosciences, 36(12), 706-716.

Sporns, O. (2014). Contributions and challenges for network models in cognitive neuroscience. Nature Neuroscience, 17(5), 652–660. https://doi.org/10.1038/nn.3690

Spratling, M. W. (2016). Predictive coding as a model of cognition. Cognitive processing, 17, 279-305.

Spratling, M. W. (2017). A review of predictive coding algorithms. Brain and cognition, 112, 92-97.

Sprevak, M. (2010). Computation, individuation and the received view on representation. Studies in History and Philosophy of Science Part A, 41(3), 260–270. https://doi.org/10.1016/j.shpsa.2010.07.008

Sprevak, M. (2024). Predictive coding I: Introduction. Philosophy Compass, 19(1), e12950.

Sprevak, M. (unpublished a). Predictive coding III: the algorithmic level. Preprint at: https://philsci-archive.pitt.edu/19488/ cited with permission

Sprevak, M. (unpublished b). Predictive coding IV: the implementation level. Preprint at: https://philsci-archive.pitt.edu/19669/ cited with permission

Sprevak, M., & Smith, R. (2023). An Introduction to Predictive Processing Models of Perception and Decision-Making. Topics in Cognitive Science.

Sterling, P., & Laughlin, S. (2015). Principles of neural design. The MIT press.

Sun, Z., & Firestone, C. (2020). Optimism and pessimism in the predictive brain. Trends in Cognitive Sciences, 24(9), 683-685.

Tani, J. (2016). Exploring robotic minds. Oxford University Press.

Tootell, R. B., Hadjikhani, N. K., Mendola, J. D., Marrett, S., & Dale, A. M. (1998). From retinotopy to recognition: fMRI in human visual cortex. Trends in cognitive sciences, 2(5), 174-183.

Uttal, W. R. (2001). The new phrenology: the limits of localizing cognitive processes in the brain. MIT Press.

Von Melchner, L., Pallas, S. L., & Sur, M. (2000). Visual behaviour mediated by retinal projections directed to the auditory pathway. Nature, 404(6780), 871-876.

Walsh, K. S., McGovern, D. P., Clark, A., & O'Connell, R. G. (2020). Evaluating the neurophysiological evidence for predictive processing as a model of perception. Annals of the new York Academy of Sciences, 1464(1), 242-268.

Westlin, C., Theriault, J. E., Katsumi, Y., Nieto-Castanon, A., Kucyi, A., Ruf, S. F., ... & Barrett, L. F. (2023). Improving the study of brain-behavior relationships by revisiting basic assumptions. Trends in cognitive sciences, 27(3), 246-257.

Yamawaki, N., Borges, K., Suter, B. A., Harris, K. D., & Shepherd, G. M. (2014). A genuine layer 4 in motor cortex with prototypical synaptic circuit connectivity. Elife, 3, e05422.