

The Predictive Brain: A Modular View of Brain and Cognitive Function?

Authors

Ines Hipolito^{1*} (inesh@uow.edu.au)

Michael D. Kirchhoff² (kirchhof@uow.edu.au)

Affiliations

1) School of Liberal Arts, Faculty of Law, Humanities and the Arts, University of Wollongong, Wollongong, Australia

2) Philosophy, Faculty of Law, Humanities and the Arts, University of Wollongong, Wollongong, Australia

* Corresponding author

Abstract - Modularity is arguably one of the most influential theses guiding research on brain and cognitive function since phrenology. This paper considers the following question: is modularity entailed by recent Bayesian models of brain and cognitive function, especially the predictive processing framework? It starts by considering three of the most well-articulated arguments for the view that modularity and predictive processing work well together. It argues that all three kinds of arguments for modularity come up short, albeit for different reasons. The analysis in this paper, although formulated in the context of predictive processing, speaks to broader issues with how to understand the relationship between functional segregation and integration and the reciprocal architecture of the predictive brain. These conclusions have implications for how to study brain and cognitive function. Specifically, when cognitive neuroscience works within an acyclic Markov decision scheme, adopted by most Bayesian models of brain and cognitive function, it may very well be methodologically misguided. This speaks to an increasing tendency within the cognitive neurosciences to emphasise recurrent and reciprocal neuronal processing captured within newly emerging dynamical causal modelling frameworks. The conclusions also suggest that functional integration is an organising principle of brain and cognitive function.

Funding – MK was funded by an Australian Research Council Discovery Project (DP170102987). IH was funded by the University of Wollongong International Scholarship.

1. Introduction

The question that will concern us is: do some of the latest and most exciting Bayesian approaches to brain and cognitive function, most notably predictive processing, support the notion that brain and cognitive function is *modular*? Modularity has had a lot of importance and influence in both neuroscience and philosophy of mind: with implications for our understanding of brain and cognitive function; the relation between different cognitive functions; and for the development of methodologies by which to conduct research in philosophy of mind and neuroscience.

Since the early anatomical theories of Gall & Surzheim (1809/2001), the notion of *modularity* has been a central research imperative in cognitive neuroscience, seeking to identify specific brain regions with particular cognitive functions. This has been called *neural localisationism* (Hohwy 2007; cf. Friston 2002). Modularity of this kind implies that the brain should be

understood as segregated into many different parts, each realising a specialised cognitive function. Neural localisationism was motivated by phrenology in the 19th Century (Gall & Surzheim 1809/2001). It has seen something of a renaissance with the advent of functional imaging techniques such as functional magnetic resonance imaging, lending support to the notion that particular regions of the brain realise specific cognitive functions - e.g., speech comprehension in the Wernicke area in the left temporal lobe. Evidence for anatomical or neural localisationism appears overwhelming. Yet appearances can be misleading. For example, Uttal (2001) makes the observation that there is still disagreement about the specific location of Broca's area - associated with language production. Evidence from lesion studies is equally problematic, if used as support for neural localisationism. Visual neglect, for instance, has been found to be associated with lesions in very different parts of the brain.

Evidence for neural localisationism is difficult to find. For this reason, modularity of this type is no longer well-supported by findings in cognitive neuroscience. This has led researchers in the field to develop and endorse functional and informational approaches to modularity, as opposed to strict neural localisationism.

In philosophy, the most celebrated notion of modularity is presented in Fodor's (1983) *The Modularity of Mind*. Fodorian modularity provides a general characterisation of modularity, such that a system counts modular if it has several features associated with modularity, to a reasonable degree (Fodor 1983, p. 37). Fodorian modularity can be cast as a combination of functional and informational modularity. On the one hand, it states that low-level perceptual and language processing is modular, whilst central processing involved in reasoning and belief-formation is non-modular. This implies functional modularity given the presence of functional segregation between modules and general processing, with modules operating in a domain-specific fashion. On the other hand, Fodorian modularity states that modules are informationally encapsulated, which implies that a module can do its work and undergo changes without affecting or being affected by other modules or higher cognitive levels of processing.

Fodorian modularity suggests that cognitive architectures consist of some modules. In this sense, it stands in contrast with neural localisationism. It should also be distinguished from a functional analogue of neural localisationism; namely, massive modularity (Carruthers 2006). This is the view that brain and cognitive function is wholly modular: from (low-level) perception to (high-level) thinking, planning and decision-making. One sees this view in the evolutionary psychology paradigm (Cosmides & Tooby 2002). At the opposite extreme, one finds holistic theories such as Lashley's (1950) equipotentiality hypothesis, which states that the brain is an undifferentiated mass. This would imply that all of the brain is involved in all kinds of cognitive functions and processing. Of course, no one is eager to defend this form of implausible holism (Prinz 2006).

In contemporary neuroscience, especially cognitive and systems neuroscience, the focus on functional specialisation, i.e., modularity, is still very much alive. Functional neuroimaging has been very successful in establishing functional segregation as a principle of organisation in the human brain (Büchel & Friston 1997). Crucially, however, is the additional observation that in these fields of brain research, functional specialisation is understood as an "extrinsic property of cortical processing that depends essentially on functionally integrated, causal relations between pairs of areas." (Hohwy 2007, p. 327) In this sense, recent advances show that functional specialisation depends on functional connectivity. As Hohwy says: "these kinds of proposals hold enough explanatory promise to go beyond blob-ology [neural localisationism] without just proposing a bland connectionism [holism]." (2007, p. 327) Some of this recent work on functional segregation and integration has been addressed within Bayesian models of brain and cognitive function; specifically, the predictive

processing (or Bayesian brain) framework, where several authors are arguing that predictive processing supports a unique mix of functional and informational modularity.

Our ultimate aim in this paper is to assess what we identify to be the three most well-articulated arguments in favour of modularity premised on predictive processing. We do not consider the massive modularity framework any further in this paper. Predictive processing, if correct, rules out this view of brain and cognitive function. Not only does it seek to provide a unified theory of perception, cognition and action (cf. Clark 2013). It also shows how there is functional integration across cognitive domains sometimes portrayed as functionally segregated (cf. Hohwy 2007). For example, attention is central to both perception and action; perception is inherently related to active sampling of the environment; and episodic memory is integrated with processes involved in planning and decision-making.

The three arguments in favour of modularity premised on predictive processing are: (a) the epistemic (Bayesian) courtroom argument; (b) the intransitivity argument; and (c) the Markov blanket argument. We briefly introduce these arguments now.

The first argument is ‘the epistemic (Bayesian) courtroom argument’ (Hohwy 2013). This argument motivates functional and informational modularity in the Fodorian sense. It asserts that perceptual inference exhibits canonical features of modularity; specifically, functional segregation and informational encapsulation. It proceeds by way of analogy to a courtroom, arguing that “[r]eality testing is somewhat like engaging an epistemic courtroom” (Hohwy, 2013, p. 152). The idea motivating the analogy is that exteroception (vision, hearing, etc.) functions akin to isolated witnesses delivering fine-grained information to perceptual inference (the judge and jury) *vis-a-vis* a specific event in question. The more sensory independent witnesses are (and the way each one independently taps into different aspects of the same event), the more efficient the process of Bayesian model optimisation is.

The second and third arguments for modularity we call ‘formal arguments,’ given their grounding in formalisms from machine learning and probability theory. We identify two arguments. The first – the *intransitivity* argument – aims to establish that causal influence across hierarchical levels in probabilistic models do not exhibit intransitivity (Drayson 2017). In this sense, the intransitivity argument underpins notions of functional and informational modularity given considerations about causal chains in Bayesian networks. The second argument turns to the Markov blanket formalism in machine learning. It relies on the notion of a Markov blanket to ground the idea that predictive processing exhibits modularity given the conditional independence and functional segregation between different hierarchical levels induced by the presence of a Markov blanket, or multiplicity of Markov blankets (Hohwy 2016). We thus call this the *Markov blanket* argument for modularity.

Each argument for modularity has its own unique and compelling features. Yet we ultimately argue they come up short, albeit for different reasons. If this is on the right track, the result is important because dominating theories of brain and cognitive function speak against what many take to be an established truth; namely, that modularity is at the explanatory basis of our best understanding of brain and cognitive function. Positively put, the results arrived at in this paper present hierarchical message passing in the brain as integrated and recurrent over multiple spatial-temporal scales. From systems neuroscience, we now know that global network dynamics arise from local dynamics, where global dynamics constrain local activity such that the entire brain becomes a self-organising system (Deco et al. 2012). Showing that there are no grounds for thinking that predictive processing yields a modular view of brain and cognitive function provides support for the hypothesis that global network dynamics enslaves activity at lower scales of brain activity. Furthermore, our results highlight that when models of brain and cognitive function is premised on an acyclic Markov decision scheme, utilised by most Bayesian models of brain and cognitive function, it is methodologically misguided. This lends once again support to the notion that the brain is a directed cyclic network, where brain and cognitive function is essentially tied to functional

integration. So, while we do not aim to revolutionise the brain and cognitive sciences in this paper, we aim to clarify what we deem to be problematic about assuming modularity in this area of research.

2. The Epistemic (Bayesian) Courtroom Argument

In this section we consider the courtroom argument (in short) for modularity. This particular argument has the notion of *conditional independence* at its base. An example will help clarify this notion. Intuitively, the observation that the room is cold could be explained either by a window having been left open; or, given that the air conditioning system having been left on some excessively high value. If you were to make the further observation that the air conditioning system is on high, then the observation that it is cold now carries no information about whether the window is open. This example shows that the coldness in the room and the window being open are conditionally independent, given that the air conditioning is on high assuming that the two events (open window, air conditioning system on high) can be treated as statistically independent. We can note, more formally, that for any variable A , A is conditionally independent of B , given another variable C if and only if the probability of A given C and the probability of B given C can be written as $p(A|C)$ and $p(B|C)$. In other words, A (it being cold) is conditionally independent of B (the window being open) given C (high air conditioning) if, when C is known, knowing A would provide no additional information about B (Beal 2003).

Conditional independence within the brain can be captured by appeal to the brain's "organic structure preventing information flow across processing streams in lower parts of the cortical hierarchy" (Hohwy 2013, p. 252). In neuroscience, this property of the brain is also known as *functional segregation*. Crucially, according to Hohwy, the chief architect of the courtroom argument, this "is not too dissimilar from the courtroom where witnesses might be kept in separate rooms, and be prevented from phoning each other." (2013, p. 252) Hence, the courtroom argument, if correct, implies that brain and cognitive function can be captured in terms of functional (segregation) modularity and informational modularity given the presence of conditional independencies between lower and higher levels with the *predictive* structure of the central nervous system. This conclusion is reached by appeal to the assumption that predictive processing – i.e. prediction error minimisation routines – can be understood by analogy to how a verdict is reached in court.

The motivating premise of predictive processing is that the brain is able to keep its states within a limited and bounded set of states, under the assumption that the brain is a locally ergodic system. In other words, it can on average and over time avoid states with high surprise (i.e., prediction error) by having a good or close to optimal model of its own states and its wider environment. First, the environment is cast as having a true hidden variable or state, s , which generates outcomes (i.e., observable sensory states), o . In this sense, outcomes are elicited by hidden states in the world, and these states can be further influenced by the action policies an agent pursues (Parr & Friston 2017). Crucially, hidden states are 'hidden' because the information an agent has access to is contained wholly and exhaustively in observable sensory states. The relationship between hidden states and outcomes is usefully captured by the notion of a *generative process*. Note that it is not feasible to determine the hidden states resulting in outcomes solely on the basis of the information contained in the outcomes. This has been shown to be computationally intractable (Murphy 2012), given that the relation between s and o is non-linear. Second, predictive processing solves this problem by appeal to the idea that the brain has a *generative model* of the generative process. That is, brains are able to approximate the true hidden probabilistic state causing its sensory observations by inferring the probability of different hidden states given some observation, $P(s|o)$, by leveraging its generative model, which is formally defined as a joint probability distribution between a prior belief, $P(s)$, and a

likelihood function, $P(o|s)$. If successful, the brain will encounter no or only minimal prediction error signals, i.e., deviations between its predictions about the causes of outcomes and the actual outcomes it has access to. Third, and finally, the accuracy and credibility of inferring the true hidden cause of a specific observation can be helped significantly by keeping different sensory modalities and higher-level predictions isolated from one another. The same, so Hohwy (2013) argues, is the case in a courtroom, where a “defendant’s claim is being interrogated in the light of supporting or contradicting witnesses” (p. 152). Accordingly, to ensure the accuracy and credibility of a verdict, witnesses, prosecutor and defence, should be segregated from one another. To illustrate this, we quote Hohwy at length:

It is essential for a trial that different witnesses are independent. This is why the witnesses should not be allowed to chat about the case in the corridor or meet with the defendant before being let in to testify. If the witnesses are not independent with respect to the case in question, then we cannot trust their evidence. Similarly, the judge needs to be impartial and the jury needs to be independent too. For example, the jury members should not be bigots and should not have had access to copious media reports before the trial begins; and the judge is not allowed to influence any testimony. If the judge and jury are not independent in this sense then we cannot trust them to evaluate and weigh the evidence in a fair way. This imposes a kind of evidential architecture on the legal system: checks and balances are in place to ensure independence. Fairness is violated when this fails. (2013, p. 152)

The suggestion is that Bayesian inference over neural states maps onto an evidential architecture akin to that of the courtroom. Likewise, different “sources of evidence need to be independent witnesses with respect to the event in question, and higher-level expectation needs to evaluate evidence from lower levels in a balanced way.” (2013, p. 152) The reason being that if evidence is influenced by multiple sensory modalities at the same time, this lowers the credibility of making use of it as a means to testing one’s predictions up and against the world (Hohwy 2013, p. 152).

Translate this into the language of modularity and you get the following picture. Processes of Bayesian model optimisation (i.e., updating the generative model in light of new evidence) entail functional segregation and informational encapsulation in the specific sense of Fodorian modularity. Hohwy is explicit about this as he says: “different parts of the system specialize in different tasks and seem relatively unimpeded by other processes (this relates to Fodor’s (1983) notion of informationally encapsulated cognitive modules).” (2013, p. 152) Lower and higher levels of the Bayesian hierarchy are taken to be functionally segregated from each other, on the one hand, and informationally encapsulated given the presence of conditional independence between levels, on the other. Hohwy (2013) calls to this ‘horizontal insulation’. He treats the relationship between sensory modalities in a similar fashion, which he refers to as ‘vertical insulation’ (Hohwy 2013).

Returning to courtroom analogy. Sensory states (witnesses) act as channels for testimonial evidence, which are arguably horizontally segregated (i.e., epistemically and functionally segregated) from higher-level cognitive states. Higher-level states, like any good judge sitting at the top of the hierarchy, should be impartial and carefully consider incoming information. Further substantiation for horizontal insulation comes from the idea that Bayesian model optimisation can be cast as a form of inference to the best explanation. In inference to the best explanation, when a hypothesis hi best accounts for — i.e. explains away — some occurring evidence ei , the latter becomes evidence for the former in so far as hi accounts for ei . In this sense, hi becomes self-evidencing. As Hohwy puts it:

When hi is self-evidencing, there is an explanatory-evidentiary circle (EE-circle) where hi explains ei and ei in turn is evidence for hi . In Bayesian terms, generative models — when inverted — generate predictions (hypotheses) that explain away prediction error (the sensory evidence), thus maximizing its evidence.” (2016, p. 263).

The presence of EE-circles is argued to induce a form of epistemic seclusion conditioned on conditional independence between different levels across the evidentiary (Bayesian) architecture, implying vertical insulation of sensory modalities, on the one hand, and horizontal insulation between sensory modalities and higher-level operations (i.e., the judge), on the other.

The question to be considered now is: how well does the courtroom analogy work? In what follows, we consider examples raising doubts about both horizontal and vertical insulation, thereby questioning the entailment relation from predictive processing to both functional and informational modularity. In sub-section 2.1., we target cross-modal predictive processing, focusing on vertical insulation between exteroception, interoception and proprioception. In sub-section 2.2, we rely on recent work on so-called embodied predictive processing that runs counter to the idea of horizontal insulation in the domain of the classical exteroceptive senses.

2.1. Cross-modal predictive processing: against vertical insulation

The first item of Hohwy’s prediction-driven account of modularity is vertical insulation, i.e., functional and informational modularity between exteroceptive sensory modalities. The courtroom argument states that exteroceptive sensory modalities are informationally encapsulated from one another. Evidence for this might seem easy to come by. We do not hear with our noses, after all. This might make vertical insulation appear almost inevitable. There is however evidence for vertical insulation being difficult to come by. Here we draw on empirical work from the literature on brain connectivity and multisensory integration to press on the results of the courtroom argument.

Multisensory integration is the view that sight, sound, smell, taste, and touch are integrated to influence perception and action (Stein et al., 2009; Stein, 2012). It refers to the neural processes by which unisensory signals are combined to enhance disambiguation. Findings show that exteroceptive sensory modalities modulate each other during different but overlapping stages of neural activity (Ball et al. 2018; Ferraro et al. 2019). If coherent, this raises doubt about the vertical insulation between exteroceptive sensory modalities. Atilgan et al. (2018), for example, show that visual stimuli shape how auditory cortical neurons respond to sound mixtures, by eliciting changes in the phase of the local field potential in auditory cortex (see Fig. 1).

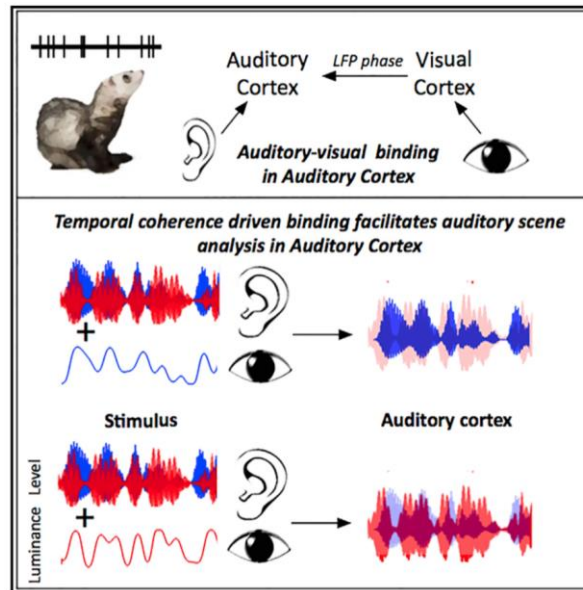


Figure 1 Temporal coherence between auditory and visual stimuli shapes the sound scene in auditory cortex. The auditory cortex is identified as a site of multisensory binding, with inputs from visual cortex underpinning these effects (Atilgan et al. 2018).

This makes it hard to support informational encapsulation in exteroception. Indeed, Atilgan et al. (2018) observed that vision effects are lost when the auditory cortex is reversibly deactivated. This means that sensory signals not only affect one another but that activation of one sensory modality depends on the activation of another sensory modality (Schroeder & Foxe 2004). Findings like these make the courtroom argument hard to defend.

The classic (exteroceptive) five senses are of course only part of the story about how agents sense their world and themselves. Under predictive processing, interoception (internal states of the body) and proprioception (the location of the body and its potential for action) play key roles in the overall process of prediction error minimisation. For example, motor governed actions such as plucking an apple from a tree comprise not only exteroceptive sight and touch but also proprioceptive sensations of location and trajectories of bodily parts. Proprioception influences exteroception as visual sensing interacts with eye and head movement (Parr and Friston 2018). Interoceptive sensations such as hunger influence exteroception, given that these kinds of internal sensations motivate action and make salient things in the environment for sight and touch, say. To see this a little more clearly, consider that your hunger would hardly be satisfied by the accuracy and credibility of an exteroceptive verdict. Motor-system governed biomechanical action requires perception, in the sense that interoception is what motivates the action in the first place (hunger). However, it also requires proprioception given that action is part of what enables organisms to minimise prediction error - e.g., via eye movement and touch (Linson et al. 2018). As Linson et al. put it: “we must recognize the continuous interactions between interoception and the other modalities ... [given that] the generative model embodies continuous relationships between extero-, proprio-, and interoception” (2018, p. 3). All modalities – exteroception, interoception to proprioception – mutually influence one another in the overall quest to minimise prediction error.

2.2. Cross-modal predictive processing: against horizontal insulation

Evidence of cross-modal predictive processing is being extended under the notion of embodied predictive processing (Allen et al. 2019; Allen and Friston 2018; Allen et al. 2016;

Limanowski and Blankenburg 2013; Wilkinson et al. 2019). The idea is that sensory evidence in hierarchical predictive processing is dependent on and influenced by bodily processes. This is motivated by empirical evidence suggesting the presence of coupling and modulation between bodily cycles (biorhythms) and neuronal oscillations and behaviour (Herrero et al., 2017; Tort et al., 2018), which brings into question horizontal insulation.

Allen et al., (2019) provide a proof of principle of this via computer simulations. They focus on the coupling between interoception and exteroception – what they call multimodal integration – mapping interoceptive cardiac cycles to exteroceptive stimuli. They do so by simulating a cardiac arousal response to threatening stimuli (seeing a vicious looking spider) in comparison to non-arousing stimuli (seeing flowers) within a Markov decision process (MDP) scheme. An MDP determines the transitions between states in generative models of the dynamics of the (hidden) variables, constituted by hidden states and actions or policies. Generative models embody the combinations between these variables. An observation is generated from the hidden states. States are generated by a time transition, depending on the state at the previous time, and on the policy (Fig. 2).

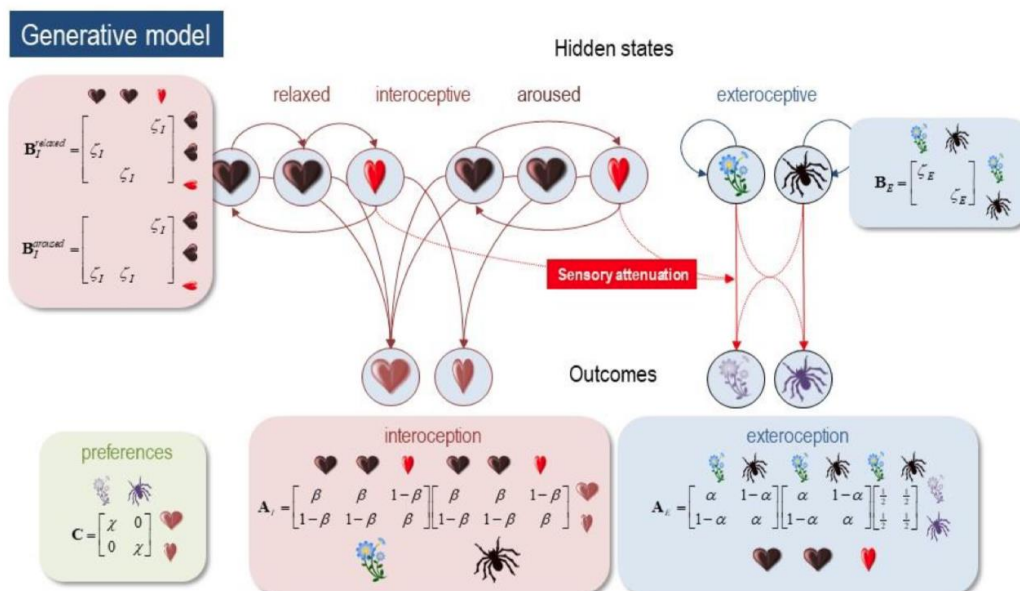


Figure 2 This schematic illustrates how hidden states cause each other and sensory outcomes in the interoceptive and exteroceptive domain. The upper row describes the probability transitions among hidden states (seeing a flower or seeing a spider), while the lower row specifies the outcomes that would be generated by combinations of hidden states that are inferred on the basis of outcomes. (Allen et al., 2019, p. 8).

Outcomes are generated according to preferences (green box) that couple exteroception (flower; spider) with interoception (diastolic state; systolic state). Outcomes depend upon the multimodal integration of the policies (i.e., being relaxed conditioned on seeing a flower; conversely, being aroused given spider observations). Probabilistically, if the subject is aroused (hidden states on the top row) the likelihood of seeing a spider is high (outcomes on the bottom row). This means that it is the interoceptive state that ensures the precision on the mapping from hidden states to outcomes.

Though interoceptive and exteroceptive outcomes are portrayed separately, they causally influence one another. According to preferences, a spider causes diastole, and diastole indicates that a spider was seen, in which case, arousal is the proper response (i.e., the one affording the least expected prediction error). Conversely, a flower causes systole and systole indicate that a flower was seen, in which case relaxing is the appropriate effect.

This means that sensory outcomes in the interoceptive and exteroceptive domain are caused by the combination of these hidden states. The result of the simulations is that exteroception is dependent, in a probabilistic sense, on interoception. In other words, one appraisal or prediction about the world (vicious spider, beautiful flower) turns on inferring hidden states (arousal, relaxed) via interoceptive processing. The generative model maps the probability over hidden states, outcomes, and policies for multimodal integration of intero- and exteroception.

This runs counter to the claim about horizontal insulation made in the courtroom argument, as interoception and exteroception are coupled to one another in a hierarchical sense, from the body to brain, and back again. Of course, those sympathetic to the courtroom argument might try to press the following point: namely, that not all information is relevant to the judge; and by extension, not all information is relevant to the brain. In this sense, modularity follows given that the brain is informationally selective. There is no problem with making this specific observation. It is correct to say that not all kinds of information are relevant at any given point to neural processing. Yet this observation does not warrant any further claims about informational encapsulation, which the courtroom argument relies on. Hence, the courtroom argument is guilty of this kind of fallacious interference: from informational selectivity to informational encapsulation.

3. The Intransitivity Argument

This argument starts from the observation that prediction error minimisation unfolds across a cortical hierarchy. On this view, priors at each level in the hierarchy are provided by the level above and prediction errors from the level below. That is, in hierarchical message passing schemes such as predictive processing, prediction error signals are influenced by states at the same level and by states at the level above. Conversely, predictions are influenced by states at the same level and by states at the level below. Under Bayesian model optimisation, to invert a generative model (i.e., to approximate the posterior probability of a prediction in light of new evidence) requires input from the level below in the form of a prediction error signal (Friston 2009). This means that hierarchical message passing “drive expectations ... towards better predictions to explain away prediction error.” (Friston 2009, p. 297) Crucially, hierarchical message passing implies that connections between predictions and prediction errors are mutual or reciprocal. Or, put differently, the only activities linking different hierarchical levels are *forward* prediction errors and *backward* mediating predictions. This speaks to why hierarchical predictive processing supports the idea that forward and backward connectivity is pervasive in brain and cognitive function (Friston 2009; Hohwy 2007).

The intransitivity argument is intended to block this conclusion; to show that this species of large-scale mutual influence is not the case in the predictive architecture of the brain. The argument is due to Drayson (2017), and can be given the following form. *Premise one*: If predictive processing is non-modular, then activity across different hierarchical levels causally influence one another. Credibility for premise one comes from the principle of transitivity. In particular, if “Level $A + 1$ influences Level A , and Level A influences Level $A - 1$, then Level $A + 1$ influences Level $A - 1$.” (Drayson 2017, p. 8) *Premise two*: activity across different hierarchical levels do not causally influence one another. *Conclusion*, predictive processing is modular. Drayson (2017) states that Bayesian networks such as predictive processing implement a form of non-transitive computation. They exhibit or instantiate “mechanisms that implement *causal* Bayesian networks.” (Drayson 2017, p. 9) Crucially, according to Drayson, probabilistic causal models are not models where causal influence is likely to be preserved over long causal chains. She takes this to suggest that the “further apart the levels in the [predictive] hierarchy are, the less likely there is to be causal influence from the higher level to the lower level.” (2017, p. 9)

The intransitivity argument is developed within the formalism of Bayesian networks. Yet it is an argument that ultimately pitches the *causal* architecture of predictive processing in terms of ‘directed acyclic graphs’ (DAGs), as opposed to ‘directed cyclic graphs or models (DCMs)’. A DAG is a form of structural causal modelling that picks out conditional independence between nodes in a network. For example, in figure 3, X_3 is statistically independent of X_1 given X_2 . This means it is possible to understand the properties and dynamics of X_3 independent of X_1 . DCMs, on the other hand, capture causal influence between nodes standing in reciprocal relations in a network.

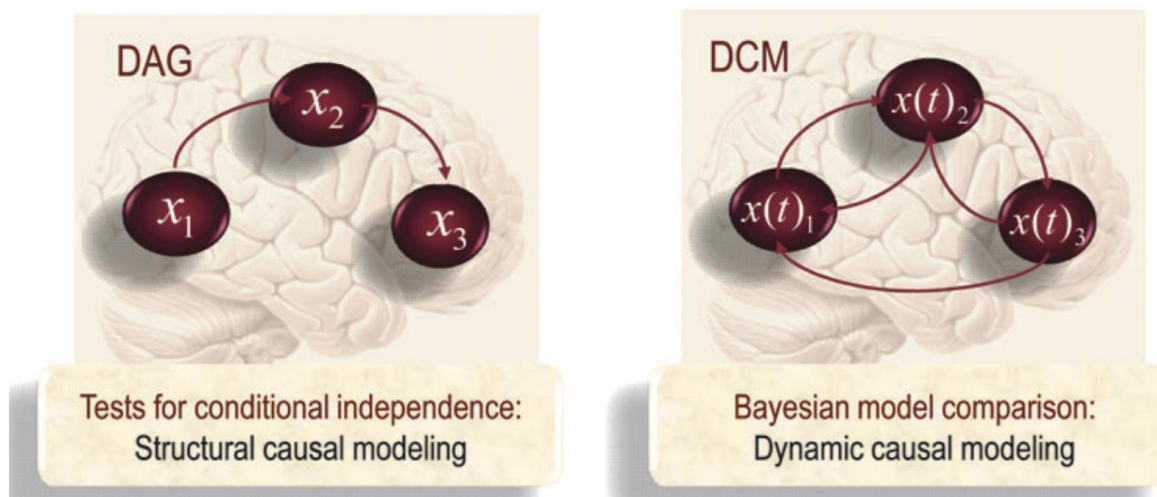


Figure 3 Schematic highlighting the distinction between structural causal modelling in the form of a directed acyclic graph and dynamic causal modelling in the form of a directed cyclic model (Friston 2011, p. 25)

The argument we shall present turns on the distinction between DAGs and DCMs. We shall argue that the intransitivity argument is based on Bayesian (causal) computation understood in terms of DAGs. We start by motivating why the intransitivity argument is formulated in terms of DAGs. This will involve some degree of formality. Yet the formalities are important, as they show why the argument under consideration concludes as it does. We then go on to highlight two limitations with using DAGs to discover causal connectivity in Bayesian networks such as the brain. We conclude that DCMs, having none of the limitations of DAGs, are consistent with predictive processing and do not imply modularity. In fact, DCMs capture the importance of functional and effective connectivity to understand brain and cognitive function in cognitive neuroscience.

To foreshadow the structure of our argument: DAGs are often used in computational neuroscience to model dependencies between states; yet this way of modelling the brain is a methodological simplification; it is not intended to show that state dependencies are in fact acyclic. Moreover, claiming that a particular system was in a specific state at a particular point in time, is actually to say that the average of the system’s states was in that particular state during that period of time. As Spivey has pointed out: “This kind of coarse averaging measurement is often a practical necessity in science, but should not be mistaken as genuine evidence for the system actually resting in a discrete stable state.” 2007, pp. 30-31). Crucially, the intransitivity argument conflates a way of modelling the brain with what is actually the case about neural-based state dependencies. This is the key problem for the intransitivity argument: it conflates models with the actual thing we want to understand. We spell this out in further detail below.

A DAG is also known as a Bayesian network (Pearl 1988). Bayesian networks are sometimes referred to as causal networks. The intransitivity argument helps itself to this notion. Yet there is nothing inherently causal about directed graphical models, of which a DAG is an example (Murphy 2012). The key property that merits attention is that a DAG has a particular topological ordering. This ordering implies that the nodes comprising a DAG can be ordered such that *parents* come prior to *children*, which over the course of time become parents, and so on. This ordering is also known as an *ordered Markov property*. Crucially, for any network, if that network has an ordered Markov property, then a node only depends on its immediate parents, and consequently not on all other parents in the topology of the network. A DAG is thus an acyclic graph without any directed cycles between nodes. For this reason, DAGs are sometimes said to be ‘memoryless’ or ‘ahistorical’, given that dependence between states of a DAG is restricted to a chain of successive influences with no reciprocal influences or loops.

This maps onto the notion of a Bayesian network assumed in the intransitivity argument. To see this in more detail, consider next that DAGs highlight one widespread way of casting decision-making and action selection in terms of a *Markov decision process* (MDP). Under predictive processing, MDPs are refined and augmented in order to deal with the only partial observability of hidden (external) causes of sensory observations. Decision-making and action selection in predictive processing are therefore more commonly approached through the lens of partially observed Markov decision processes (POMDPs). This is an important distinction in machine learning. It is however not important for present purposes, given that both MDPs and POMDPs turn on the Markov assumption that nodes stand in a relation of unidirectional conditional dependence, i.e., any given node in an MDP is governed wholly and exhaustively by its immediately preceding states (its parent). Modelling causal pathways in predictive processing via DAGs therefore fits snugly with the intransitivity argument, as it breaks with the idea of large-scale causal connectivity between multiple nodes in a network.

All of this is standard machine learning textbook. Nevertheless, there are two problems with utilising DAGs to argue for functional and informational modularity in Bayesian models such as predictive processing. The first is that it restricts conditional independence between nodes in a network to the strict and linear topological ordering of a Markov property. As Friston has noted, this “is problematic because the brain is a directed cyclic graph [or model]. Every brain region is connected reciprocally (at least polysyntactically), and every computational theory of brain function rests on some form of reciprocal or reentrant message passing.” (2011, p. 25) Put differently, a DAG is just a special case of a system with a Markov topology, implying that the particular properties of a DAG do not generalise to all systems with an ordered Markov property. Biological systems such as the brain can be cast with a Markov topology even if functional segregation in the brain rests on integration. The second is that DAGs ignore time; or, more accurately, it treats time as discrete, whereas time at the scale of brain dynamics is continuous.

Dynamical causal modelling (DCM), by contrast, is a generic Bayesian framework for inferring hidden (neuronal) states from brain activity (Stephan et al. 2010). Unlike DAGs, DCMs capture (1) how a particular state of some population of neuronal dynamics cause changes in other populations via synaptic connections, and (2) how these interactions change and shift given influence of external causes (e.g., endogenous brain activity). This speaks to the idea that DCM describes how distinct dynamics and regions are coupled to one another, endowing “the system with memory such that future states are influenced by current states” (Stephan et al. 2010, p. 3099). This is important for a couple of reasons. The first is that DCMs underpin a central tenet of predictive processing; namely, that the brain is a locally ergodic system. That is, the recurrent dynamics of DCM captures the notion that the states that an agent (i.e., nervous system) occupies are bounded and limited, and that the

dynamics operate so as to enable the agent to keep frequenting those states, again and again, given that it is this limited distribution of states that define the agent as the kind of organism that it is. The second is that DCMs avoids being 'memoryless' and 'ahistorical' - something DAGs are vulnerable to. Neural dynamics are self-organising and adaptive, and therefore regulate their changes with respect to certain viability constraints. DAGs capture this by constraining the temporal extension of such processes such that state changes in a system always unfold at a particular point in time as a result of current dynamics. DCMs however illustrate how past changes inform future states given the current state of the system in question. This allows DCMs to capture how memory is essential to cognitive functions such as planning and decision-making. Finally, DAGs represent state transitions in discrete time. Natural systems such as brains however function in continuous time. Real or continuous time "does not function like a digital computer's clock. It does not move forward and then stop to be counted, and then move forward again only to stop again." (Spivey 2007, p. 31) DCM thus provides a more 'natural' way of modelling state transitioning in hierarchical predictive processing.

More specifically, DCMs allow for the representation of hierarchical message passing in the brain that takes the form of a *directed cyclic graph* (Fig. 4).

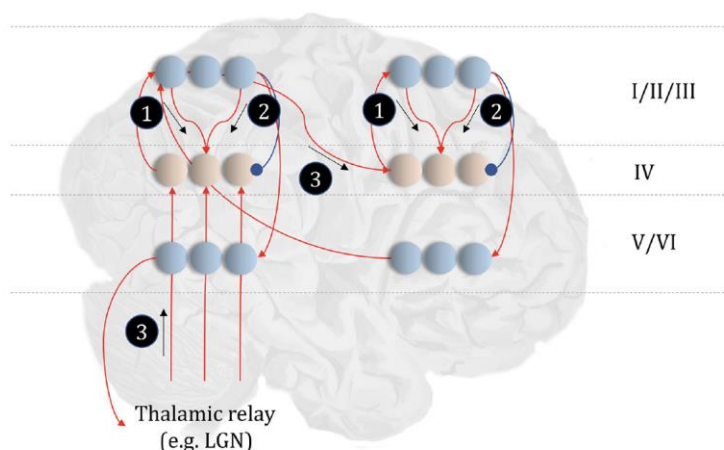


Figure 4 Schematic of neurobiological instantiation of perceptual inference in predictive processing (Parr & Friston 2018, p. 6)

This figure is a general model of the neurobiological instantiation of perceptual inference in which perceptual inference takes the form of hierarchical message passing, represented as a directed cyclic graph. Unlike a DAG, a directed cyclic graph defines a series of nodes with recurrent and reciprocal connections in a network or system, which nevertheless respects the Markov topology crucial to modelling state transitions in terms of POMPDs. In figure 4, layer IV represents nodes (spiny stellate cells) influenced by lower-level nodes (relay nuclei) of the thalamus, and from lower cortical levels (e.g., the lateral geniculate nucleus). Layer IV nodes signal prediction errors, reflecting a mismatch between layer V/VI predictions and sensory input received via the thalamic relays. If this is sound, it suggests that perceptual inference is best cast in the form of hierarchical message passing represented by a directed cyclic graph; not a DAG. This is important, for it speaks against the conclusion of the intransitivity argument. The intransitivity argument states that there are no large-scale reciprocal causal connections across the inferential or predictive hierarchy. DCM informed analyses of perceptual inference suggest otherwise. Crucially, although figure 4 only represents reciprocal message passing across a limited number of levels, this pattern could be recursively extended to any arbitrary number of levels. Specifically, in their discussion of figure 4, Parr & Friston (2018) go on to conclude that perceptual inference rests on anatomical and functional interconnectivity over long causal chains. Crucially, our argument

does not jeopardise the notion that state transitions within DCM schemes exhibit an ordered Markov topology; it undercuts the idea that the brain and its transitions can be understood as intransitive across levels of organisation.

We wish to finish this section by considering a possible objection; namely, that while it might be true that modelling hierarchical message passing in terms of directed cyclic graphs preserves transitivity across levels, a treatment of the same issue in the form of DAGs still supports the intransitivity argument. Here is the problem should one pursue this kind of response. It makes the argument for modularity *model relative*. This is not an outcome an argument for modularity should (even implicitly) accommodate, for the notion of modularity is only really informative if it gets at the actual architecture of brain and cognitive function. Let us therefore look at a different line of support for modularity in predictive processing.

4. The Markov Blanket Argument

Previously we considered the role of conditional independence in the courtroom argument for modularity, and found it wanting. A different and more applicable route from formalism to biology, when aiming to secure the functional segregation claim, is likely to come from the Markov blanket formalism, and how it works in hierarchical message passing. We call this the Markov blanket argument for modularity. It elaborates on the notion that different levels in the perceptual hierarchy stand in a relation of conditional independence. The Markov blanket argument makes this a centrepiece in an argument for modularity in predictive processing. So, we now turn to conditional independence in the context of Markov blankets. Like the intransitivity argument, the Markov blanket argument focuses on the notion of horizontal insulation between lower and higher levels in perceptual inference. Before we consider the Markov blanket argument, we start by motivating the notion of a Markov blanket.

Pearl (1998) introduced the Markov blanket into machine learning and probability theory. In his treatment, it was used to denote a set of properties specific to Bayesian networks (like DAGs and DCMs). Figure 5 provides a general schematic of the Markov blanket for a node.

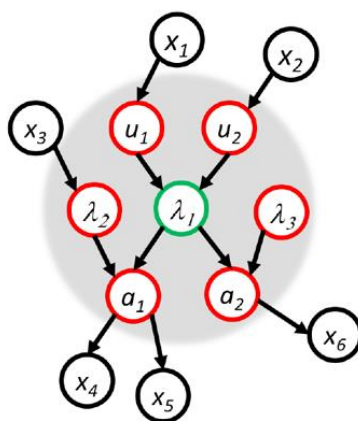


Figure 5 Schematic depiction of a Markov blanket with full conditionals (Hohwy 2017, p. 3).

The Markov blanket is a statistical boundary, comprised of parents, children and parents of children. It renders internal states, green node, conditionally independent from external states, black nodes. This means that if one is trying to predict the behaviour of a state with a Markov blanket, then knowing the states comprising the Markov blanket, red nodes, will be sufficient. This implies that external states are rendered uninformative with respect to predicting the values of Markov blanketed states. It is this kind of statistical blanket for the

green node in figure 5 that is referred to as a Markov blanket. Technically, a Markov blanket can then be defined as the parents (u), the children (a), and the parents of the children (s).

In biology, more specifically, the notion of a Markov blanket allows one to define any system in a way that delineates it from the environment in which it exists. In this sense, the Markov blanket is best understood as a boundary (Schrödinger 1943). Markov blanket boundaries can be found at many different scales of life, from macromolecules to organelles, organs and humans (Clark 2017; Hipólito 2019; Kirchhoff et al. 2018; Kirchhoff & Kiverstein 2019; Palacios et al. 2017; Ramstead et al. 2017). A key aspect of a Markov blanket is that it yields a formal way by which to define what it means for internal and external states to be conditionally independent of one another given a third set of states: active and sensory states. Specifically, the partitioning rule of a Markov blanket states that internal states can only influence external states via active states, while external states can only influence internal states via sensory states. Thus, the partitioning rule of the Markov blanket formalism precludes *direct coupling* between internal and external states. Crucially, the delineation between states only arises given coupling. As a cell, for example, separates itself from its extracellular milieu, it remains statistically coupled to it given the states that define its Markov blanket (Fig. 6).

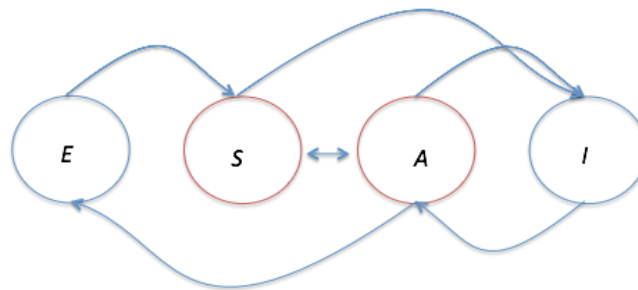


Figure 6 Illustration of the partitioning rule governing Markov blankets (Kirchhoff & Kiverstein, 2019)

This figure highlights the conditional independencies induced by the presence of a Markov blanket. On the one hand, external states, E , cause sensory states, S , which influence, but are not also influenced by, internal states, I . On the other hand, internal states cause active states, A , which influence, but are not themselves influenced by, external states (Friston et al. 2015; Kirchhoff et al. 2018; Palacios et al. 2017).

The question we now turn to is: how does the Markov blanket play a role in making a case for modularity in predictive processing, and is it justified? We shall argue that even the appeal to the Markov blanket formalism comes up short in an argument for modularity; in particular, in an argument for horizontal insulation. Before we present our argument, let us first present the Markov blanket argument for modularity.

The argument is due to Hohwy (2013). The form of the argument can be cast as follows. *Premise one:* given the presence of a Markov blanket (aka EE-circle), it becomes possible to draw a principled distinction between internal causes as they are inferred by a model (i.e., predictions) and external causes, outside of the Markov blanket. Crucially, this holds over the entire predictive hierarchy given that it is possible to distinguish between higher-level predictions, lower-level input and intermediate error signals. On this view, then, any level in the predictive hierarchy has its own Markov blanket, i.e., sensory and active states. This means that internal states (the generative model) can influence the immediate level below via top-down predictions, and external states at a lower level can influence higher levels via sensory signals (prediction errors). Hence, strictly speaking, for any level in hierarchical message passing this division between states holds. *Premise two:* the statistical division of states induced by a Markov blanket is sufficient for evidential insulation. As Hohwy says:

The perceptual hierarchy, and the iterations of prediction error minimization mechanisms implemented at each of its levels, seem apt to deliver evidential insulation between horizontal levels. Each overlapping pair of levels forms a functional unit where the lower level passes prediction errors to the higher level, and the higher level passes predictions to the lower level ... In this sense the upper level in each pair of levels only “knows” its own expectations and is only told how these expectations are wrong, and is never told directly what the level below “knows” ... For this reason the right kind of horizontal insulation comes naturally with the hierarchy. (2013, p. 153)

The *conclusion* is that brain and cognitive function exhibit horizontal insulation. This is the Markov blanket argument for modularity. A few remarks about this argument before moving on to develop an alternative to modularity premised on the Markov blanket formalism. It is formulated epistemically, in terms of what higher and lower levels ‘know’ and ‘do not know’. This epistemological formulation need not be the right formulation. Others are possible (Bruineberg et al. 2018). Regardless, it signals the evidential seclusion needed to arrive at a conclusion about horizontal insulation. It does this by utilising the idea that a Markov blanket around a node or level shields such a level (epistemically, statistically, or some other variation) from any other node or level, and vice-versa. So, the argument for modularity is formulated with an eye to conditional independence between nodes (or states at different levels of hierarchical organisation).

We now develop a reason for resisting the conclusion of the Markov blanket argument, yet without resisting the Markov blanket formalism. One can start to develop an answer to why Markov blankets do not imply modularity by taking into consideration initially that the statistical structure of the Markov blanket formalism is *scale-free*. That it is scale-free implies that it allows for the formation of Markov blankets at recursively larger and larger scales of organisation. It is not restricted to any particular scale of systemic organisation. This means that at any level of the predictive hierarchy one can define a Markov blanket consisting of active and sensory states that separate higher and lower levels, resulting in a view of Markov blankets comprised of multiple other Markov blankets (Fig. 7).

Blankets all the way down

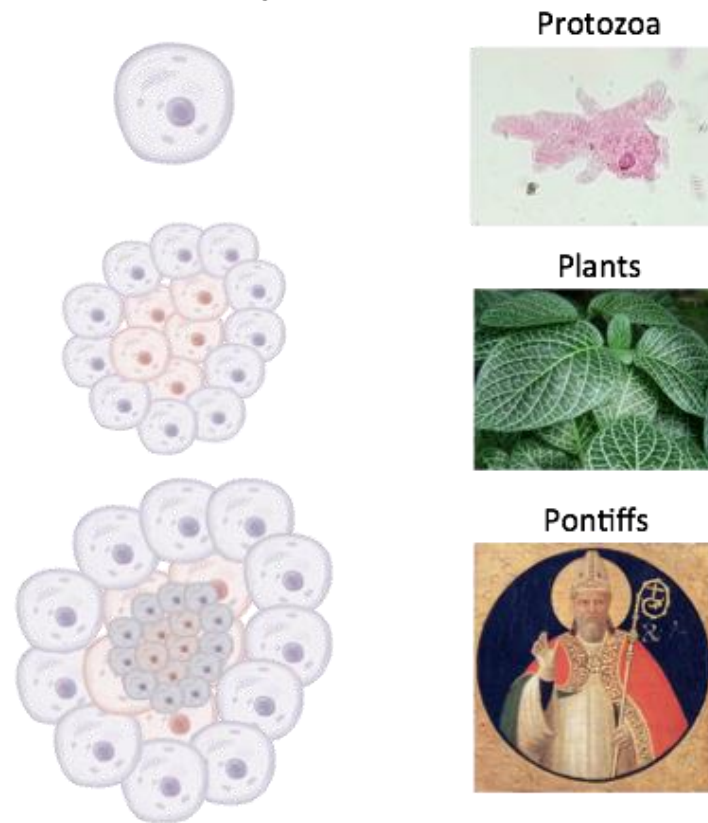


Figure 7 Schematic depiction of Markov blankets. The top figure depicts a single Markov blanket. The middle figure represents a multiscale and nested organisation of Markov blankets. The final figure suggests that cultural practices can envelope a multiplicity of individuals given its nested structure. Thus figure 7 represents the Markov blanket organisation all the way down to individual cells and all the way up to complex organisms like human beings (Kirchhoff et al. 2018; cf. Clark 2017).

To see how this scale-free organisation of Markov blankets speaks against the notion of horizontal insulation in predictive processing, we need to consider the reciprocal or circular causal interactions that occur across the hierarchy of Markov blankets. First, it is a common observation that low levels of the hierarchy predict causal regularities unfolding at very fast, millisecond, time-scales, whereas more complex regularities unfold at higher levels and over much slower time-scales. This suggests that prediction errors are minimised over the entire hierarchy at any given time. Second, Hobson & Friston (2014) have argued that macroscale dynamics constrain microscale dynamics in hierarchical predictive processing. A different way of putting this would be in terms of multilayered Markov blankets such that Markov blankets at the microscale give rise to Markov blankets at the macroscale, while Markov blankets at the level of the ensemble constrain the activities of Markov blankets nodes at lower and lower levels in the predictive hierarchy. This is entirely consistent with the idea of slower unfolding dynamics at higher levels and faster evolving dynamics at lower levels in the hierarchy. This division of timescales between the micro- and macro-scale in dynamical systems such as the brain is a signature feature of the slaving principle in statistical physics and synergetics (Haken 1983; Hobson & Friston 2014). The slaving principle shows how activities at the microscale are constrained by activities at the macroscale such that activities at the microscale no longer behave independently of activities at the macroscale but are “sucked into an ordered coordinated pattern.” (Kelso 1995, p. 8) This is entirely consistent with work in the cognitive neuroscience focus on global brain dynamics. As Deco et al. (2012) put it:

Global network dynamics over distributed brain areas emerge from the local dynamics of each brain area. Conversely, global dynamics constrain local activity such that the whole system becomes self-organizing. The implicit coupling between local and global scales induces a form of circular causality that is characteristic of complex, coupled systems that show self-organization, like the brain. (2012, p. 18)

Palacios et al. (2017) have developed a proof of concept of this idea via computer simulations, showing the self-organisation of an ensemble Markov blanket, comprised of fifteen Markov blankets within it, with the Markov blanket at the global level constraining the dynamics of every Markov blanket at the local level (see Palacios et al. 2017 for detailed discussion).

So here is the real problem with using the Markov blanket formalism to run an argument for horizontal insulation. If the Markov blanket formalism does not imply insulation of this kind, it cannot be used in support of modularity. It does not follow from this that the presence of a scale-free and multilayered Markov blanket organisation does away entirely with the notion of functional segregation. For lower and higher levels in the hierarchy differ in their processing style, at least with respect to the timescales over which they unfold. Modularity however does not come for free, even once this much is acknowledged. The reason for this is that functional segregation, given the Markov blanket formalism, can only be understood in the context of functional integration (Friston 2011). This is because the segregation between states induced by Markov blankets only arise given the presence of integration or coupling. Or, put differently, a state or node differentiates itself from another state or node, yet remains statistically coupled to it. This follows because two states or more are conditionally independent not merely due to spatial distance or separation, but given the states comprising the Markov blanket: active and sensory states.

There may be one way for a defender of modularity in predictive processing to attempt to save modularity. It is easy enough to imagine that defenders of the intransitivity argument and the Markov blanket argument might take this route. They may wish to give up on the philosophically substantial notion of modularity from Fodor (1983), adopting instead a much weaker notion. Integrative modularity provides a weakening of Fodorian modularity. It picks out the idea of functional segregation, at least to a certain extent. There are two problems with opting for a weak notion of modularity. The first we have already stated: in hierarchical message passing, functional segregation is premised on functional integration. This is not likely to be a satisfying result for the proponent of modularity. The second turns on value in the context of explanation. Choosing to adopt a weak notion of modularity has a price; it threatens to render the notion itself explanatorily vacuous. Specifically, if the notion ceases to exhibit any of its robust properties, enabling it to make a difference to description or explanation, then what is the value of keeping it? It is not evident that there is any value.

5. Conclusion

This paper has considered arguments for modularity in the context of predictive processing and active inference. Three kinds of arguments for modularity were reviewed: (a) the Bayesian courtroom argument; (b) the intransitivity argument; and (c), the Markov blanket argument. We argued that each of these pro-modularity arguments has its own compelling features, although all three of them come up short in showing that predictive processing entails modularity in the context of brain and cognitive function. More importantly, our conclusions present hierarchical message passing in the brain as integrated and recurrent over multiple spatial-temporal scales. Furthermore, we argued that when models of brain and cognitive function is premised on an acyclic Markov decision scheme, they will tend to be misguided. We showed how this supports the idea that the brain is a directed cyclic network, where brain and cognitive function is essentially tied to functional integration. Our

conclusions therefore speak to how to understand brain and cognitive function in the absence of taking modularity for granted.

References

Allen, M., Levy, A., Parr, T., & Friston, K. J. (2019). In the Body's Eye: The Computational Anatomy of Interoceptive Inference. *BioRxiv*, 603928.

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

Allen, M., Fardo, F., Dietz, M. J., Hillebrandt, H., Friston, K. J., Rees, G., et al. (2016). Anterior insula coordinates hierarchical processing of tactile mismatch responses. *NeuroImage*, 127, 34–43.

Atilgan, H., Town, S. M., Wood, K. C., Jones, G. P., Maddox, R. K., Lee, A. K., & Bizley, J. K. (2018). Integration of visual information in auditory cortex promotes auditory scene analysis through multisensory binding. *Neuron*, 97(3), 640-655.

Ball, F., Michels, L. E., Thiele, C., & Noesselt, T. (2018). The role of multisensory interplay in enabling temporal expectations. *Cognition*, 170, 130-146.

Beal, M.J. (2003). *Variational Algorithms for Approximate Bayesian Inference*. PhD Thesis, University College London, UK.

Büchel, C., and Friston, K. (1997). Modulation of connectivity in visual pathways by attention: cortical interactions evaluated with structural equation modelling and fMRI. *Cerebral Cortex*, 7(8), 768-778.

Bruineberg, J., Kiverstein, J., & Rietveld, E. (2018). The anticipating brain is not a scientist: The free-energy principle from an ecological-enactive perspective. *Synthese*, 195, 2417–2444.

Clark, A. (2017). How to knit your own Markov blanket: Resisting the second law with metamorphic minds. In T. Metzinger and W. Wiese (Eds.), *Philosophy and Predictive Processing: 3. Frankfurt am Main: MIND Group*. Doi: 10.15502/9783958573031.

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

Cosmides, L., & Tooby, J. (2002). Unraveling the enigma of human intelligence: Evolutionary psychology and the multimodular mind. *The evolution of intelligence*, 145-198.

Deco, G., Jirsa, V., & Friston, K. J. (2012). The dynamical structural basis of brain activity. *Principles of brain dynamics: Global state interactions*, 1.

Drayson, Z. (2017). Modularity and the predictive mind. In T. Metzinger & W. Wiese (Eds.). *Philosophy and Predictive Processing: 12. Frankfurt am Main: MIND Group*. doi: 10.15502/9783958573130

Ferraro, S., Van Ackeren, M. J., Mai, R., Tassi, L., Cardinale, F., Nigri, A., ... & Collignon, O. (2019). Stereotactic electroencephalography in humans reveals multisensory signal in early visual and auditory cortices. *bioRxiv*, 549733.

- Fodor, J. A. (1983). *The modularity of mind*. MIT press.
- Friston, K., Rigoli, F., Ognibene, D., Mathys, C., Fitzgerald, T., & Pezzulo, G. (2015). Active inference and epistemic value. *Cognitive neuroscience*, 6(4), 187-214.
- Friston, K. J. (2011). Functional and effective connectivity: a review. *Brain connectivity*, 1(1), 13-36.
- Friston, K. (2009). The free-energy principle: a rough guide to the brain?. *Trends in cognitive sciences*, 13(7), 293-301.
- Friston, K. (2002). Beyond phrenology: what can neuroimaging tell us about distributed circuitry?. *Annual review of neuroscience*, 25(1), 221-250.
- Gall, F. J., Spurzheim, J. K., & Oehler-Klein, S. (1809/2001). *Untersuchungen ueber die Anatomie des Nervensystems ueberhaupt, und des Gehirns insbesondere*. Georg Olms Verlag.
- Haken, H. (1983). Synopsis and introduction. In *Synergetics of the Brain* (pp. 3-25). Springer, Berlin, Heidelberg.
- Herrero, A., Duhart, J. M., & Ceriani, M. F. (2017). Neuronal and glial clocks underlying structural remodeling of pacemaker neurons in drosophila. *Frontiers in physiology*, 8, 918.
- Hipólito, I. (2019) A Simple Theory of Every 'Thing'. *Physics of Life Reviews*.
- Hobson, J. A., & Friston, K. (2014). Consciousness, dreams, and inference: the cartesian theatre revisited. *Journal of Consciousness Studies*, 21(1-2), 6-32.
- Hohwy, J. (2017). How to entrain your evil demon. In T. Metzinger and W. Wiese (Eds.), *Philosophy and Predictive Processing: 3. Frankfurt am Main: MIND Group*. Doi: 10.15502/9783958573048.
- Hohwy, J. (2016). The self-evidencing brain. *Noûs*, 50(2), 259-285.
- Hohwy, J. (2013). *The predictive mind*. Oxford University Press.
- Hohwy, J. (2007). Functional Integration and the mind. *Synthese*, 159(3), 315-328.
- Kelso, J. S. (1995). *Dynamic patterns: The self-organization of brain and behavior*. MIT press.
- Kirchhoff, M. D., & Kiverstein, J. (2019). *Extended Consciousness and Predictive Processing: A Third Wave View*. Routledge.
- Kirchhoff, M., Parr, T., Palacios, E., Friston, K., & Kiverstein, J. (2018). The Markov blankets of life: autonomy, active inference and the free energy principle. *Journal of The royal society interface*, 15(138), 20170792.
- Kirchhoff, M.D. (2018). Predictive brains and embodied, enactive cognition: an introduction to the special issue. *Synthese*. Volume 195, Issue 6, pp. 2355–2366.
- Lashley, K. (1950). In search of the engram. *Symposia of the Society for Experimental Biology*, 4, 454-482.

Limanowski, J., & Blankenburg, F. (2013). Minimal self-models and the free energy principle. *Frontiers in Human Neuroscience*, 7.

Linson, A., Clark, A., Ramamoorthy, S., & Friston, K. (2018). The active inference approach to ecological perception: general information dynamics for natural and artificial embodied cognition. *Frontiers in Robotics and AI*, 5, 21.

Murphy, K. P. (2012). *Machine learning: a probabilistic perspective*. MIT press.

Palacios, E. R., Razi, A., Parr, T., Kirchhoff, M., & Friston, K. (2017). Biological Self-organisation and Markov blankets. *BioRxiv*, 227181.

Parr, T., & Friston, K. J. (2017). The active construction of the visual world. *Neuropsychologia*, 104, 92-101.

Parr, T., & Friston, K. J. (2018). Active inference and the anatomy of oculomotion. *Neuropsychologia*, 111, 334-343.

Pearl, J. (1988/2014). *Probabilistic reasoning in intelligent systems: networks of plausible inference*. Morgan Kaufmann Publishers.

Prinz, J. (2006). Is the mind really modular? In R. Stainton (ed.), *Contemporary Debates in Cognitive Science* (pp. 22-36). Blackwell.

Ramstead, M. J. D., Badcock, P. B., & Friston, K. J. (2017). Answering Schrödinger's question: a free-energy formulation. *Physics of life reviews*, 24, 1-16.

Schroeder, C. E., & Foxe, J. J. (2004). 18 Multisensory Convergence in Early Cortical Processing. *The handbook of multisensory processes*, 295.

Schrödinger, E. (1943). *What is life?*, Based on lectures delivered under the auspices of the Dublin Institute for Advanced Studies at Trinity College, Dublin.

Stein, B. E. (Ed.). (2012). *The new handbook of multisensory processing*. Mit Press.

Stein, B. E., Stanford, T. R., & Rowland, B. A. (2009). The neural basis of multisensory integration in the midbrain: its organization and maturation. *Hearing research*, 258(1-2), 4-15.

Tort, A. B., Brankač, J., & Draguhn, A. (2018). Respiration-entrained brain rhythms are global but often overlooked. *Trends in neurosciences*, 41(4), 186-197.

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

Wilkinson, S., Deane, G., Nave, K., & Clark, A. (2019). Getting Warmer: Predictive Processing and the Nature of Emotion. In *The Value of Emotions for Knowledge* (pp. 101-119). Palgrave Macmillan, Cham.

Word count: 9714