

Special issue of *Synthese*, guest-edited by Jakob Hohwy

With contributions by Mike Anderson, Rick Grush, Chris Eliasmith, Karl Friston & Klass Stephan, Tori McGeer, Tim Bayne & Elisabeth Pacherie, Phil Gerrans.

Published online, *Synthese*, Oct 2007.

Functional integration and the mind

Jakob Hohwy

Monash University

Introduction

Across the diverse subdisciplines of the mind sciences, there is a growing awareness that, in order to properly understand the brain as the basis of the mind, our theorising and experimental work must to a higher degree incorporate the fact that activity in the brain, at the level of individual neurons, over neural populations, to cortical areas, is characterised by massive interconnectivity.

In brain imaging science, this represents a move away from ‘blob-ology’, with a focus on local areas of activity, and towards functional integration, with a focus on the functional and dynamic causal relations between areas of activity. Similarly, in theoretical neuroscience there is a renewed focus on the computational significance of the interaction between bottom-up and top-down neural signals.

At the cognitive, psychological and philosophical levels, there is also more focus on functional integration, for example concerning interactions between volition and representation; between agency and movement; between the self and representation; between experience, inference and belief formation; and between representations of self and other, as well as language and communication.

This emerging integrative approach suggests, for example, that in order to understand belief formation and agency we must also understand visual representation, and vice versa. Or, that to understand speech comprehension we must understand the motor control system. Or, that to understand attention and self we must understand what happens in the resting state. Or, that in order to understand psychosis we need to understand volition and the sensorimotor system as well as reasoning. In short, that in order to understand the significance of one “blob” we need to understand the significance of many blobs and their interactions.

This special issue of *Synthese* explores this trend in the mind sciences from a broadly philosophical perspective. The question is: what does a more integrative approach to brain function and cognition tell us about the nature of the mind; in particular, what should it make us say about topics such as representation, inference, belief formation, agency, self, emotion, as well as psychopathology? Will functional integration challenge previous conceptions of brain function in new and unexpected ways? The contributors are well-versed in interdisciplinary neuroscience and bring each their perspective and expertise to bear on these questions.

In this introduction, I explain the motivation for this special issue in more detail and I introduce a particular theoretical framework that has the potential to unify much of the discussion on functional integration. En route, I relate the central issue to the individual contributions to this issue (though, I hasten to say, of course without committing all the authors to the unifying framework and without being able to capture all the rich detail of their individual articles).

Learning from blobs: the route to functional integration.

Neuroscience, in particular, cognitive and systems neuroscience, should teach us about the nature of mind and cognition. It should explain the mind. The question is, how can this be achieved?

A major tool for neuroscience these years is brain mapping. Brain mapping is the activity of using, for example, functional Magnetic Resonance Imaging (fMRI) in localising areas of activity in the brain associated with various cognitive and motor tasks. An example is some widely reported studies by Bartels & Zeki (see 2004) locating the neural substrates of love. Subjects are scanned while presented with pictures of loved ones, and with various closely matched pictures for control, and the difference in activation in these different conditions suggest overlapping areas of activity for romantic and maternal love in the striatum, the insula and dorsal anterior cingulate cortex; combined with deactivations focused at the middle prefrontal cortex, the parieto-occipital junction, the medial prefrontal/paracingulate cortex and the temporal poles). Thousands of such studies, of a very wide range of cognitive functions, have been published in the last decade or so.

On their own, such localisation studies do not tell us much about the mind, as Karl Friston reminds in a paper entitled “Beyond phrenology” (2002: 221). We do not get any wiser about the nature of love from being told that its neural substrate includes the anterior cingulate cortex. We can illustrate this by imagining what would happen had it been found that love instead correlates with activity in the dorsolateral prefrontal cortex. It wouldn’t, on its own, teach us anything different about the nature of love.

One could conclude from this that imaging neuroscience is rather pointless (for a comprehensive criticism of brain imaging, see Uttal 2001). But such a conclusion would be hasty. Firstly, such studies can be of clinical importance; for example, in predicting what kinds of deficits patients with organic damage to these areas are likely to experience. Notice though that the accuracy of such predictions depends on how closely one can map functions to brain areas, an issue that we touch on below. Secondly, these imaging studies very rarely can, or should, be taken on their own. A whole body of different studies may be taken into consideration when discussing the significance of findings such as the one for maternal and romantic love. As Bartels & Zeki discuss, it turns out that some of the deactivated areas are normally “associated with negative emotions, social judgment and ‘mentalizing’, that is, the assessment of other people’s intentions and emotions.” (2004: 1155). This brings them to conclude that:

human attachment employs a push– pull mechanism that overcomes social distance by deactivating networks used for critical social assessment and negative emotions, while it bonds individuals through the involvement of the reward circuitry, explaining the power of love to motivate and exhilarate (2004: 1155).

This is an intriguing suggestion about the nature of maternal and romantic love, namely that it is associated with lowering the standards when it comes to loved ones: it is not that we see their faults and try to explain them away, it is that we simply don’t register them. It may even help

explain the excruciating misreading of some famous love object's true feelings in erotomania. This way of learning from imaging studies rests on the standard scientific method of integrating diverse findings in one's interpretation of data.

But this will only take us so far. We might in the end have a complete map of partially overlapping areas of activity for a large number of functions. This would tell us something about what functions share what properties. But it would fall short of explaining *how* the brain enables those cognitive and motor functions, that is, of explaining the mechanisms that produce such mental phenomena. The problem is that mapping doesn't tell us about what goes on in and between the mapped areas. It tells us that there is activity, not what the activity is.

In order for neuroscience to explain the mind, then, it must discern at least the nature of the brain activity associated with the cognitive and motor functions one is investigating. Naturally, one useful thing will be to investigate the neural processing that goes on at each location of activity.

However, this will not, on its own, take us far either. The reason is that it is, of course, a mistake to think that cognitive and motor functions map one-to-one onto isolated areas of activity in an otherwise inactive brain. (This is the mistake behind simplistic media reports of neuroscience findings that often make it appear as if one coloured blob superimposed on a standard brain can show what love or crime (or whatever) is). Rather, it is fundamental to brain function that each area of activity is connected to other areas, that the same area is recruited for parts of a number of different functions, and that all this interconnected activity happens in the context of relatively high levels of global brain activity. It is therefore highly unlikely that mind and cognition can be explained without recourse to neural interconnectivity and global modulatory context.

Concretely, when viewing cognitive functions in isolation from each other it is difficult to explain why the same areas may be recruited by different functions, and, when viewing areas of activation in isolation from each other it is difficult to explain the role of other areas that are also recruited by that function. Even if we look at a number of integrated functions and a number of areas simultaneously, we would need to take into consideration the modulatory effect of the brain's overall pattern of activity (Raichle 2006).

Some of these problems can be illustrated by some recent developments in understanding the role of the dorsolateral prefrontal cortex (DLPFC), as discussed by Richard Frackowiak and his London colleagues (2004):

In the early days of functional imaging every task seemed to activate DLPFC, and every experimenter was happy to define a different role for this region. For example, it was proposed that DLPFC was critical for willed action [...], for working memory [...], or for semantics [...]. The tasks used in these studies were complex and involved many processes. Inevitably the selection of one of these processes to be associated with DLPFC was somewhat arbitrary (p349).

Careful theoretical reviews, and refinements in the tasks chosen in newer experimental paradigms, allows a more principled approach to the role of DLPFC:

By identifying a series of different circumstances in which DLPFC is activated in association with response selection we have tried to derive a single cognitive function for this region. [...] We suggest that DLPFC is most likely involved in

defining a set of responses suitable for the task and biasing these for selection, when external inputs do not automatically achieve such selection. (p357).

This is progress on the earlier more arbitrary guesses at its function, and yet Frackowiak et al are not quite satisfied:

Nevertheless, the formulation we have put forward remains very crude. A major problem stems from the fact that most of our discussion has been restricted to consideration of DLPFC in isolation. Only when we can discuss in truly mechanistic terms how it interacts with other areas will we begin to gain a proper understanding of its role in executive function (p357)

To address this latter issue, it is necessary to combine investigation of functional localisation with investigation of the interactions between areas, that is, with investigation of *functional integration*.

The increased awareness of functional integration is essentially concerned with the causal dynamics of neural interconnectivity, that is with the causal relations between areas of activity, given their temporal evolution and context. From the perspective of the philosophy of neuroscience this is interesting for it coincides with a renewed interest in causal dynamic modelling at large (see, e.g., Pearl 2000; Woodward 2003), as well as with a related interest in mechanistic explanation. It looks as if many explanations in this area will be computational and probabilistic so further issues will concern how computational theory and probabilistic approaches to brain activity can have explanatory power (see Piccinini 2006 as well as the papers in the special issue of *Synthese* 153(3) 2006, edited by Piccinini). These issues come up indirectly in the papers of this special issue of *Synthese* too. However, the main question here is what functional integration teaches us about the mind. That is, even though the study of functional integration is still only in its beginning phases, it may point us in new directions and challenge our previous conceptions about a broad range of cognitive, affective and agential mental phenomena.

Between functional localism and functional integrationism

To this point, we have noticed how brain imaging very often shows activity in a number of areas, and how different cognitive tasks often activate overlapping areas; and we have remarked on how understanding the causal interactions between areas of activity may be important. In the light of this, it will be difficult to insist that cognitive function can be wholly explained by reference to local computational solutions. In other words, what is needed is an alternative to *strong functional localism*.

A conservative alternative to strong localism is what we might call *strong functional, but anatomically distributed localism*: the view that cognitively inaccessible processing is realised in an anatomically distributed network of areas. This would begin to deal with how brain imaging finds a number of anatomically distributed activations pr task. A worry about this is that very many areas can show up, suggesting that localism really is not an appropriate way to think about brain processing. For example, in studies of binocular rivalry, there is rivalry-related activations found in temporal cortex, in early visual cortex as well as the lateral geniculate nucleus (Haynes, Deichmann et al. 2005), and activity in right prefrontal cortex is associated with the perceptual transitions (for review, see Tong, Meng et al. 2006). The emerging picture is that very many areas of the visual processing system are involved in producing, in this case, conscious perception. Given this tendency for activity to be distributed, it seems unclear at what point a believer in anatomically

distributed localism should begin to think that “localism” is no longer an apt term to describe his or her view.

Similarly, strong functional, but anatomically distributed localism does not explain why the same areas, such as the DLPFC show up in a number of different tasks. One solution would be to use brain imaging results as clues in attempts to identify subfunctions for commonsense or traditional psychological functions. The analysis of the DLPFC that Frackowiak et al attempted above is an example of such an approach. Though the proposal for the DLPFC is intriguing, much work lies ahead here, as Frackowiak et al candidly conclude:

It is our hope that the reader will have found this chapter [on executive function] highly unsatisfactory. We have considered some topics in great detail, while, at the same time, failing to discuss a number of important studies. We have drawn conclusions about the role of certain regions though careful selection of the evidence. [...]. If we are to understand the role of prefrontal cortex in higher cognitive functions we need to take into account a much wider range of evidence. We need to be much more precise in our characterization of cognitive functions. We need to be much more accurate in our localization of brain activity. (p357)

On its own, such an approach could be called *strong subfunctional localism*. As we noted, however, we will only be able to understand the role of such subfunctions if we can connect them properly with the other subfunctions that make up a commonsense or a traditional psychological function. This task is fraught with empirical and conceptual difficulties because we need a highly integrated approach that allows us to explain how the same strongly localised subfunction can play a role in different overall functions, in order to explain how it can be recruited in different such overall functions.

What is the contrast to localism? Mike Anderson, in his contribution here, makes a very clear distinction between localism and holism, based on considerations akin to those above. He argues that the contrast to localism is not a holism where one implausibly claims that all of the brain is involved in all cognitive functions. Rather, holism is a view according to which no area is devoted exclusively to one function, and where the same area contributes differently to the different functions it subserves. Anderson identifies an intermediate position according to which each area is redeployed to make the same contribution to several different functions. These are very general hypotheses about overall brain architecture and as such one should think that empirical evidence is hard to come by. Anderson however sees that the evolutionary notion of exaptation could be integral to the redeployment thesis and is able to formulate a series of predictions on this basis, such as that more phylogenetically recent functions should rest on more widely distributed neural areas than older ones. These hypotheses are tested successfully on a wide battery of fMRI studies. This approach allows us to explore more concretely what functional integration means for the mind. For example, Anderson notes how speech understanding and motor control may have deep, shared subfunctional elements that intriguingly suggests how the more recent linguistic capacities may be partly enabled by a conception of motor ‘doability’ such that a doable action indicates a comprehensible sentence. In this way studies of functional integration can very clearly point to new and unexpected avenues for understanding the nature of the mind.

Computer metaphors and functional integration

It may indeed be that a redeployment thesis midway between untenable localism and implausible holism is most promising. Such a thesis is however neutral when it comes to explaining what the mechanistic or causal contribution of each area is, *how* it works in the cognitive economy. One possibility is captured by a kind of simplified computer analogy for the mind according to which each area works on its own subroutines and only interacts with other areas by passing input and output back and forth between areas. However, there are reasons to believe that this is not a very likely picture; instead, a more truly integrationist view of cognitive processing must be found.

A useful pointer in relation to this issue is found in a landmark article by David Mumford from the early 90'ties (1992). He reminds us that in the cortex the majority of cells in any one area are pyramidal with long axons that connect them to other areas, and says:

The fact that the majority of cortical cells have inter-area projection, as opposed to exclusively intra-area projections, seems already to bear an important computational message: it means that almost nothing goes on internally in one area without this activity being transmitted to at least one other area (242).

This fact tells against the simplified computer analogy. As Mumford says:

This analogy would only make sense if the number of intra-area locally projecting neurons were an order of magnitude larger than the number of inter-area globally projecting neurons. Since this isn't the case, a different paradigm must be sought (242).

Mumford proposes a paradigm that incorporates interconnectivity in a radical departure from the locationist approach of the simplified computer analogy:

[t]he bulk of the computational work of the cortex is not carried out by one area at a time, but by the information going back and forth over reciprocal pathways connecting pairs of areas (242).

On this view, the answer to how the brain works lies in the nature of the connectivity itself. The nature of the activity in one area is not as important as the effect of that activity in other areas of the brain. Of course, this might be true but still fail to impress. Just as a few blobs of activity will not explain the mind, so a few causal relations will not either. To carry explanatory weight, an appeal to interconnectivity should come with a good theoretical interpretation of the role of causal relations in the brain. Mumford offers a compelling probabilistic approach, which has much in common with ideas first proposed by Helmholtz (1860) (though an early precursor appears to be Ibn al-Haytham (Alhacen) around 1010AD), and revived in various forms by for example Mackay (1956), Neisser (1967), and Gregory (1980). Versions of this overall approach is, as we will see, introduced, developed and explored in several papers in this special issue.

Perceptual inference, prediction, surprise and death

As Mumford insists, we need to take the massive interconnectivity of the brain into consideration. This relates in particular to the extensive network of corticocortical loops and corticothalamic loops in the brain. This prompts Karl Friston to say:

[T]he representational capacity and inherent function of any neuron, neuronal population, or cortical area in the brain is dynamic and context sensitive. Functional integration, or interactions among brain systems, that employ driving (bottom-up) and backward (top-down) connections mediate this adaptive and contextual specialization (2002: 247).

That is, in getting at the proper role for the causal interactions among cortical areas, we need to work systematically with the idea that there are bottom-up and top-down signals with different functional roles. Bottom-up would go, roughly, from sensory cortices, e.g., the early visual cortex at the back of the brain and forwards towards the prefrontal cortex. Top-down goes in the opposite direction. Friston proceeds to give a formal account and development of arguments that are “developed under generative models of brain function, where higher-level systems provide a prediction of the inputs to lower-level regions” (247). Here, the appeal to predictions is central. We can use Richard Gregory (1997) to motivate this idea:

Following von Helmholtz's lead [on perception as unconscious inference] we may say that knowledge is necessary for vision because retinal images are inherently ambiguous (for example for size, shape and distance of objects), and because many properties that are vital for behaviour cannot be signaled by the eyes, such as hardness and weight, hot or cold, edible or poisonous. For von Helmholtz, ambiguities are usually resolved, and non-visual object properties inferred, from knowledge by unconscious inductive inference from what is signalled and from knowledge of the object world. It is a small step [...] to say that perceptions are hypotheses, predicting unsensed characteristics of objects, and predicting in time, to compensate neural signaling delay (discovered by von Helmholtz in 1850), so ‘reaction time’ is generally avoided, as the present is predicted from delayed signals [...] Further time prediction frees higher animals from the tyranny of control by reflexes, to allow intelligent behaviour into anticipated futures (1997; see also his 1980).

The basic idea is very simple: the posterior probability of a given hypothesis about the causes of one's sensory input will go up if predictions of future input based on that hypothesis are correct. This could be the method by which the brain gets to represent the world in spite of the noise in the sensory channels and the context-sensitivity of the causes in the sensorium (such as, e.g., occlusion). This means that the brain doesn't try to infer the causes from the effects (the sensory input) but rather predicts the effects (the input) from a model of their causes. Often, the predictions will not be correct, so this simple picture needs to be supplemented with a notion of perceptual learning:

Conflict between the [bottom-up signal and the top-down predictions] is resolved by changes in the higher-level representations, which are driven by the ensuing error in lower regions, until the mismatch is "cancelled."(Friston 2002:247).

This idea can, as Eliasmith points out in this volume, be fleshed out in terms of a lower bound on learning through generative models and hierarchical Bayes, and is given a formal treatment in terms of free energy minimisation in Friston and Stephan's paper in this special issue. The changes in the parameters of the higher-order representations or models are driven by how close their predictions come to the actual input: how good one is at minimising surprise, given the brain's

current state. From this perspective the overall processing aim of the brain is to cancel out the bottom-up driving signal associated with sensory input. To the extent it can do that it will represent its world, to the extent it fails it will experience too much free energy, for example in the form of a phase shift, or death.

It is a counterintuitive but intriguing property of this kind of theory that the apparently driving, bottom-up signals are not really driving, and the backwards signals are not providing feedback. Rather, it is the bottom-up signals that have the function of providing *feedback* on the generative models that in turn provide the top-down predictions (Friston 2005; see also this volume). So it would be wrong to say such a system is self-supervised, and it would be wrong to say that it is supervised in the normal sense of having a representing agent (or programmer) adjusting synaptic weights. Instead, it is supervised by the world, the incoming signals keep a check on the accuracy of the generative models. This property of the system could have an impact on how we think about cognition in very general terms. For example, it is hard to refrain from tying this in with questions about binding. One can speculate that, if the main driving signal is in effect hierarchically ordered top-down predictions, then binding of different perceptual attributes (like shape, colour and motion) happens by fiat. At each level, the hypothesis about the input at the neural level immediately below may bind, say, colour and shape simply because they are the parameters of the hypothesis or model in question. If in turn the predictions about colour and shape are borne out, then lower level signals in areas specialised for colour and shape are matched and “cancelled” out (if the predictions were wrong, then the model is updated and new binding hypotheses generated). In turn, models at even higher levels will seek to explain the model that binds shape and colour, perhaps in turn binding them to motion. On this speculative view the binding problem doesn’t need a solution, it just dissolves.

Overall, this statistical approach to brain function gives a clear perspective on the role of causal interactions in the brain. It is not a matter of going with the simplified computer analogy, rather the activity in any area, especially the evolution of activity over time, only makes sense when seen in the context of its modulation of, or its being modulated by, activity in other areas. There is much more to be said about these broad kinds of models, for which I refer to the papers by Eliasmith, by Grush and by Friston & Stephan.

One important and, probably, unfashionable thing that this theory tells us about the mind is that perception is indirect. As Gregory (1997) puts this Helmholtzian notion:

For von Helmholtz, human perception is but indirectly related to objects, being inferred from fragmentary and often hardly relevant data signaled by the eyes, so requiring inferences from knowledge of the world to make sense of the sensory signals (1122).

What we perceive is the brain’s best hypothesis, as embodied in a high-level generative model, about the causes in the outer world. Friston & Stephan also relate the theory to attention and various psychophysical phenomena, as well as to aspects of neurophysiology and imaging neuroscience. Concretely, we have suggested that a core effect in many studies of consciousness, namely binocular rivalry, can be explained by predictive coding (Hohwy, Roepstorff et al. in prep.).

Chris Eliasmith, in this special issue, brings an interesting further insight to the debate. He notes that psychology has tended to focus on what the cognitive system realised by the brain does, *what*

its function is; and that neuroscience has tended to focus on describing *how* the brain realises its functions, whatever they are. Eliasmith argues that it is essential to address the theoretical issue of large-scale, biologically plausible functional integration in a unified manner, and proposes a mutually constrained marriage of his detailed Neural Engineering Framework (developed in collaboration with Charles Anderson) with statistical modelling of neural function of the sort adumbrated above. The importance of this is that the statistical models we have looked at so far need to be neurophysiologically realistic, they need to conform with the signalling properties that neurons actually have. We thereby get a strong vision of, in the words of Eliasmith's title, *How to Build a Brain*, since we will focus on both the what and the how.

A larger issue looms behind discussion of the extent to which *how* and *what* questions relate. It is the aim of cognitive neuropsychology to use behavioural evidence drawn from populations with abnormal functioning to deduce models of normal cognitive functioning. This approach sees abnormal functioning to be the result of damage to local areas of processing or to pathways between such areas. By examining the right kinds of deficits we can deduce how the boxes and arrows diagrams for normal functioning should look. As Tori McGeer points out in her paper, this picture downplays both the extent of redeployment and the extent of functional integration. On this perspective, no immediate explanatory role is found for neuroscience disciplines, such as functional imaging. That is, strong believers in the methods of cognitive neuropsychology believe the what and how questions are separate and not able to inform each other. McGeer argues that this view is misguided both theoretically and empirically. However, she sidelines the issue of whether someone, who believes the brain realises cognitive function, should abolish cognitive theories that cannot be realised by the brain. Instead she focuses on the extent to which neuroscience may inform hypotheses about cognitive function before realisation questions are asked, in particular she shows that interpretation of behavioural data can be, and is, laden with, neuroscientific theory. In a detailed discussion of the developmental disorder, William's Syndrome, McGeer shows that distal neuroscientific evidence about abnormal development should play a role in the interpretation of behavioural evidence. This highlights how different patterns of functional integration, for example resulting from different developmental paths, can result in similar behavioural patterns; and it demonstrates that theoretical approaches biased towards functional localism may struggle to explain the mind.

Before ending this section, we should revisit the issue of localism vs integrationism. A theory based on generative models, as the one mentioned above, does not have the consequence that we should abandon localism altogether in favour of functional integrationism. There is still functional specialization, though as Friston (2002) notes:

From this perspective the specialization of any region is determined both by bottom-up driving inputs and by top-down predictions. Specialization is therefore not an intrinsic property of any region but depends on both forward and backward connections with other areas. Because the latter have access to the context in which the inputs are generated, they are in a position to modulate the selectivity or specialization of lower areas (247).

So, what it counsels is a combination of localisation and integration, with the understanding that one cannot be understood without the other. This is in turn consistent with the views of Eliasmith and McGeer.

Perception, agency, reward.

Perceptual inference, as von Helmholtz and his followers see it, is essentially an active process. The predictions about future sensory input must be tested for inference and learning to occur, and testing requires actively sampling the environment. Predictions should therefore be functionally integrated with behaviour and agency. At its most general, sampling of the environment happens in order to maximise the posterior probability of one's hypotheses about the world, or, what amounts to the same thing, minimising perceptual surprise.

Rick Grush, in his contribution, goes deeper into the relationship between perception and behaviour. In a development of earlier work, he discusses in detail how the spatial content of perceptual experience is linked with behavioural dispositions, and explains and defends this link by providing an account of the neural processing that underlies spatial content and tying this in to emulation theory, which is of the same family of theories as the ones based on generative models and predictive coding that we mentioned above. His appeal to basis functions and the way they relate to behavioural dispositions allows him to explain why someone could be able to predict sensory input without grasping its content. Grush usefully combines all this with an intuitive illustration of what could happen when someone goes from having the spatial information available but fails to extract the spatial content or purport from that information. What we learn is that functional integration needs to be wider than just, for example, the various cortical areas of just the visual processing stream; in addition neural systems underlying behavioural dispositions must be included. On this kind of view, the cognitive system maintains internal representations through dispositions for active sampling of the environment. This agrees with the influential "enactive" approach to perception (Noë 2004) in focusing on agency but disagrees strongly with that approach in insisting on internal representation. Grush shows in detail how his approach differs from orthodox "enactivism".

We learn then that perception is imbued with behavioural dispositions, with some kind of readiness for acting on, and sampling, the world. We might further ask about the sense of agency itself. Does functional integration provide novel insights into that mental phenomenon? Tim Bayne & Elisabeth Pacherie discuss two competing approaches to the sense of agency. One approach focuses on what is sometimes called the narrative self: a holistic, domain-general module that assigns agency to the actions that fit into an overall narrative. The other approach focuses on more atomistic domain-specific comparator systems. These latter comparator systems are of a piece with the more general computational approach we have discussed above. They have been endorsed in an explanation of the sense of agency in health and in schizophrenic delusion of alien control (Frith 2005). The basic idea is that a generative model of one's motor commands allows subpersonal prediction of the sensory consequences of one's movements. If one can predict those consequences then the movement is probably one's own, if they cannot be predicted, then not. Bayne & Pacherie argue that neither approach is satisfactory on its own, in other words that to account for the sense of agency for something as simple as an arm movement we need to integrate the forward modelling comparator systems with very high level, domain-general narrative efforts.

A further, interesting question arises here. We have encountered a general theory of brain function employing generative models, and seen that forward modelling of bodily movement fits nicely what that kind of theory. Now Bayne & Pacherie argue that we need to take a narrative module into account too. The question arises whether the narrative self can be accounted for by appeal to generative models? Hohwy (2007) speculates that since generative models are kinds of

narratives, it may be that the narrative self is embodied as a generative model of a subject's desires and intentions. This view is tentatively combined with findings of global brain function and the so-called default areas of de-activation in attention-demanding tasks discovered by Raichle and colleagues (Raichle, MacLeod et al. 2001; Gusnard 2005). On this view, we represent the inner world in just the same way as we represent the outer world, which in fact should make us less surprised at the functional integration between areas concerned with desire and intention, and areas concerned with representation of the external world.

We have yet to discuss the role of affect and emotion. After having been practically shunned in neuroscience for a long time, there has, in large part due to the work of Antonio Damasio (Damasio 1994), been a surge of interest in how emotional states interact with reasoning. On our simplistic computer analogy, emotion and reason should interact by passing input and output around. This is not too far from the influential theory of emotions as somatic markers for representations of various action outcomes. When outcomes are not so marked, as may happen in patients with lesions to their ventromedial prefrontal cortex (VMPFC), practical reasoning becomes impaired. Philip Gerrans, in his contribution, agrees that there is functional integration between reasoning and emotion but compellingly shows that, independently of Damasio style positive and negative emotional valence, motivation is matter of how dopamine systems contribute to predict reward. The somatic marker theory cannot account for this kind of finding. Part of the upshot is that, though emotional valence naturally plays a role in learning, decision-making comes apart from valence. On this basis, Gerrans proposes a different account of the deficits seen in the VMPFC lesioned patients. This account appeals to the notion of mental time travel (MMT), the ability to access and use information from previous experience and imaginatively rehearse future experiences as part of the process of deliberation.

Perhaps one good way to describe MMT is in terms of relatively high-level generative models, or emulations, that are tied to decision making. On the basis of previous experience we try to find out which decisions it will be rewarding to act on by counterfactually imagining ourselves in one or the other situation. On this perspective, we should consider generative models enriched with reward parameters. So again, it is possible to take this kind of proposal in the direction of the kind of computational theory we have discussed above.

Perspectives and speculations

The papers in this special issue explore many different aspects and consequences of functional integration. They do not line up behind one uniform account of functional integration. I have attempted to indicate how some chosen aspects of the papers may fit in to a general picture of the mind that builds on unconscious perceptual inference via generative models. Hopefully this has shown that this kind of very abstract model, which has functional integration as its centrepiece, holds some promise for explaining core aspects of mind and cognition.

At the very least, we have seen how there is functional integration across domains that are often thought to be functionally segregated. Perception is essentially tied to behavioural dispositions and active sampling of the environment; behaviour and agency in turn are tied in equal measure to low-level re-afferent processing and intention and desire; decision-making is based on affect for learning and reward prediction for inference, and is enshrined in high-level imaginary projections in to the future. I also speculated that it could change the way we perceive the binding problem.

Functional integration also seems consistent with the neuroscientific evidence of neural interconnectivity, it sits well with evolutionary evidence, and makes sense of how developmental evidence can be used in cognitive science. It can be combined with very basic theories about the computational properties of neurons, the entities that brains are built of. Together, these pieces of the puzzle suggest that when traditional cognitive science works with simplified computer analogies of the mind, it may very well be methodologically misguided. Functional specialisation, modularity, is only an extrinsic property of cortical processing, it depends essentially on functionally integrated, causal relations between pairs of areas. At the same time, these kinds of proposals hold enough explanatory promise to go beyond blob-ology without just proposing a bland connectionism.

Inevitably, one thinks about how functional integration and generative models in the brain could address the problem of consciousness. It would of course be foolish to speculate too much about this here. One thing that can be said is that if perception is based on unconscious inference, then conscious content is not directly in touch with its external world; though it is supervised by the world the content or phenomenology itself is tied to the parameters of generative models. At our most ambitious, we can turn again to Gregory who makes a valiant and intriguing attempt at situating the role of consciousness in relation to the idea of perception as driven by unconscious predictive inference. As we saw, he subscribes to the idea that “[p]erceptions are like predictive hypotheses of science. Both depend on knowledge: stored data, generalisations, and assumptions.” (1996: 377). But then he goes on to suggest “that consciousness serves to flag the present.” He explains:

[h]ypotheses are essentially timeless; but the present is uniquely important for survival, as real-time decisions are essential for dealing with reality. Some features of the present are signaled by the senses, bottom-up. So the senses may not only select stored knowledge for hypotheses of what is out there; they may also signal the present for action – flagged with consciousness indicating this is for real (1996: 379).

In our terms, one may guess that the hypothesis that best predicts, and thus most efficiently cancels out, the sensory signal is the one that gets to flag the present. Gregory is perceptive and candid about the conceptual status of his proposal and says “This notion that normally consciousness flags the present *does not begin to explain* how conscious states are produced by brain processes” (379). The proposal, quite in line with many of the papers in this special issue, uses a compelling interpretation of the idea of functional integration to say something about the function of consciousness.

Acknowledgements: I wish to thank John Bickle for advice and assistance, and all the many reviewers that volunteered their time to review the papers, and, of course, the authors for contributing such a stellar collection of papers.

Bartels, A. and S. Zeki (2004). "The neural correlates of maternal and romantic love." NeuroImage 21(3): 1155-1166.

Damasio, A. (1994). *Descartes' Error*.

Frackowiak, R. S., Ed. (2004). Human Brain Function. London, Academic Press.

- Friston, K. (2002). "Beyond phrenology: What Can Neuroimaging Tell Us About Distributed Circuitry?" Annual Review of Neuroscience **25**(1): 221-250.
- Friston, K. J. (2005). "A theory of cortical responses." Philosophical Transactions: Biological Sciences **369**(1456): 815 - 836.
- Frith, C. (2005). "The self in action: Lessons from delusions of control." Consciousness and Cognition **14**(4): 752.
- Gregory, R. L. (1980). "Perceptions as hypotheses." Phil. Trans. R. Soc. Lond., Series B, Biological Sciences **290**(1038): 181-197.
- Gregory, R. L. (1996). "What do qualia do." Perception **25**(4): 377-379.
- Gregory, R. L. (1997). "Knowledge in perception and illusion." Philosophical Transactions of the Royal Society B: Biological Sciences **352**(1358): 1121-1127.
- Gusnard, D. A. (2005). "Being a self: Considerations from functional imaging." Consciousness and Cognition **14**(4): 679.
- Haynes, J. D., R. Deichmann, et al. (2005). "Eye-specific effects of binocular rivalry in the human lateral geniculate nucleus." Nature **438**: 496.
- Helmholtz, H. v. (1860). Treatise on Physiological Optics. New York, Dover.
- Hohwy, J. (2007). "The sense of self in the phenomenology of agency and perception." Psyche **13**(1).
- Hohwy, J., A. Roepstorff, et al. (in prep.). "Predictive coding and binocular rivalry."
- MacKay, D. M. (1956). The epistemological problem for automata. Automata studies. C. E. Shannon and J. McCarthy. Princeton, Princeton University Press: 235–251.
- Mumford, D. (1992). "On the computational architecture of the neocortex." Biological Cybernetics **66**(3): 241.
- Neisser, U. (1967). Cognitive psychology. New York, Appleton-Century-Crofts.
- Noë, A. (2004). Action in Perception. Cambridge, Mass., MIT Press.
- Pearl, J. (2000). Causality. Cambridge, Cambridge University Press.
- Piccinini, G. (2006). "Computational explanation in neuroscience." Synthese **153**(3): 343-353.
- Raichle, M. E. (2006). "The Brain's Dark Energy." Science **314**(5803): 1249-1250.
- Raichle, M. E., A. M. MacLeod, et al. (2001). "A default mode of brain function." PNAS **98**(2): 676-682.
- Tong, F., M. Meng, et al. (2006). "Neural bases of binocular rivalry." Trends in Cognitive Sciences **10**(11): 502.
- Uttal, W. R. (2001). The New Phrenology. Cambridge, Mass., MIT Press.
- Woodward, J. (2003). Making Things Happen. New York, Oxford University Press.