Psychology Press
Taylor & Francis Group

Modules and brain mapping

Karl J. Friston and Cathy J. Price

The Wellcome Trust Centre for Neuroimaging, University College London, London, UK

This review highlights the key role of modularity and the additive factors method in functional neuroimaging. Our focus is on structure–function mappings in the human brain and how these are disclosed by brain mapping. We describe how modularity of processing (and possibly processes) was a key point of reference for establishing functional segregation as a principle of brain organization. Furthermore, modularity plays a crucial role when trying to characterize distributed brain responses in terms of functional integration or coupling among brain areas. We consider additive factors logic and how it helped to shape the design and interpretation of studies at the inception of brain mapping, with a special focus on factorial designs. We look at factorial designs in activation experiments and in the context of lesion–deficit mapping. In both cases, the presence or absence of interactions among various experimental factors has proven essential in understanding the context-sensitive nature of distributed but modular processing and discerning the nature of (potentially degenerate) structure–function relationships in cognitive neuroscience.

Keywords: Additive factors; Modularity; Factorial; Connectivity; Degeneracy.

This review is essentially a narrative about how some of the fundaments of experimental design and interpretation of human brain mapping studies have developed over the past two decades. Its focus is on the role of modularity and additive factors logic in guiding these developments. This is a somewhat self-referential account, which allows us to describe how our earlier misconceptions gave way to more enduring perspectives—perspectives that help guide our current research into structure–function relationships in the brain.

This review comprises four sections. The first considers, briefly, the rationale for modularity and its place within distributed neuronal processing architectures. We consider evolutionary imperatives for modularity and then a slightly more abstract treatment that underpins the analysis of functional and effective connectivity. The second section is a short historical perspective that covers the rise and fall of cognitive subtraction and the emergence of factorial designs in neuroimaging. Our focus here is on the role of additive factors logic and the connection to conjunction analyses in neuroimaging. The third section pursues the importance of interactions in factorial designs—specifically, their role in disclosing context-sensitive interactions or coupling among modular brain areas. We illustrate this using the notion of dynamic diaschisis and psychophysiological interactions. The final section turns to

Correspondence should be addressed to Karl J. Friston, The Wellcome Trust Centre for Neuroimaging, Institute of Neurology, Queen Square, London WC1N 3BG, UK. (E-mail: k.friston@fil.ion.ucl.ac.uk).

This work was funded by the Wellcome Trust. We would like to thank Marcia Bennett for help preparing this manuscript.

http://www.psypress.com/cogneuropsychology
http://dx.doi.org/10.1080/02643294.2011.558835

lesion–deficit mapping and neuropsychology (in the sense of using lesions to infer functional architectures). Here, we review the concept of necessary and sufficient brain systems for a given task and how these led to the appreciation of degenerate structure–function mappings. Additive factors logic again plays a key role but, in this instance, the combination rule (Sternberg, 2011 this issue) becomes probabilistic and acquires a multiplicative aspect. We rehearse the importance of degenerate mappings in the context of multilesion–deficit analysis and conclude with some comments on the role of cognitive ontologies in making the most of neuroimaging data.

In defence of modularity

Most neurobiologists who are sensitive to the distributed and self-organized nature of neuronal dynamics tend to distance themselves from functionalist accounts of modularity. However, there is a growing appreciation of the importance of modularity in network theory (e.g., Bullmore & Sporns, 2009; He & Evans, 2010; Valencia et al., 2009) and the study of complex systems in general. The importance of modularity is usually cast in terms of robustness and evolvability in an evolutionary setting (e.g., Calabretta, 2007; Redies & Puelles, 2001). At first glance, robustness might appear to limit the evolvability of biological networks, because it reduces the number of genetic variations that are expressed phenotypically (and upon which natural selection can act; Sporns, 2010). However, neutral mutations, which are expressed in a phenotypically neutral way, can promote evolution by creating systems that are genetically varied but function equally well (i.e., degenerate, many-to-one mappings between genotype and functional phenotype). In brief, "robustness implies that many mutations are neutral and such neutrality fosters innovation" (Wagner, 2005 p. 1773). Both robustness and evolvability are enhanced by the modular organization of biological systems—from gene and protein networks to complex processes in development (e.g., Duffau, 2006). The dissociability or decomposability afforded by modularity is characteristic of the brain's small world connectivity architecture, a feature that has received increasing empirical support from analyses of anatomical and functional connectivity (Bullmore & Sporns, 2009). In short, modularity may be an emergent characteristic of any biological network that has been optimized by selective pressure (irrespective of the particular constraints on adaptive fitness). Interestingly, these arguments about modularity have emerged in a field one might least expect—namely, network theory.

The role of network theory (and graph theory) is also central to the way that imaging neuroscience tries to assess functional brain architectures. In brief, the brain appears to adhere to two principles of organization: functional segregation and integration. Functional segregation refers to the specialization of brain regions for a particular cognitive or sensorimotor function, where that function is anatomically segregated within a cortical area or system. Functional integration refers to the coupling and interactions (message passing) among these areas (Friston, Frith, Liddle, & Frackowiak, 1993). The mathematical description of networks like the brain often appeals to graph theory, where the interactions among regions (nodes) are encoded by connections (edges). These connections imply a conditional dependency between the (usually hidden) states of each region. In the brain, these hidden states correspond to population or ensemble neuronal activity performing computations. The key point here is that, to understand the network, one has to identify where there are no connections. This may sound paradoxical but emphasizes the importance of conditional independencies. Conditional independence means that knowing the activity of one area tells you nothing about the activity of a second area, given the activity in all other areas. If this can be shown statistically, one can infer the absence of a connection between the two areas in question. This absence endows the graph with a sparsity structure and a specific sort of architecture. The very existence of conditional independencies induces modularity and becomes the ultimate aim of network and effective connectivity analyses in neuroimaging.

The importance of conditional independence is reflected in the first sentence of Sternberg (2011 this issue): “The first step in one approach to understanding a complex process is to attempt to divide into modules; parts that are independent in some sense.” Mathematically, this “sense” is a conditional independence.

A key example of a connectivity analysis is dynamic causal modelling (DCM), in which one is trying to explain observed neuronal responses in terms of an underlying Bayesian dependency graph (Friston, Harrison, & Penny, 2003). In DCM, the dependencies are modelled in terms of the effective connectivity between hidden neuronal states in each brain area. Model selection is then used to identify the architecture that best explains the systems response, using Bayesian model selection. The very fact that one characterizes distributed responses in terms of a set of connected regions (nodes) speaks to the implicit modularity of processing within each node. See Figure 1 (and Seghier & Price, 2010) for an example of dynamic causal modelling in addressing the modular but distributed architectures underlying reading and object naming. It should be noted, however, that the dependencies can themselves be context sensitive. In other words, being modular does not mean responding to the same inputs in the same way all the time. Understanding this context sensitivity is one of the most important aspects of network and causal modelling, especially in cognitive

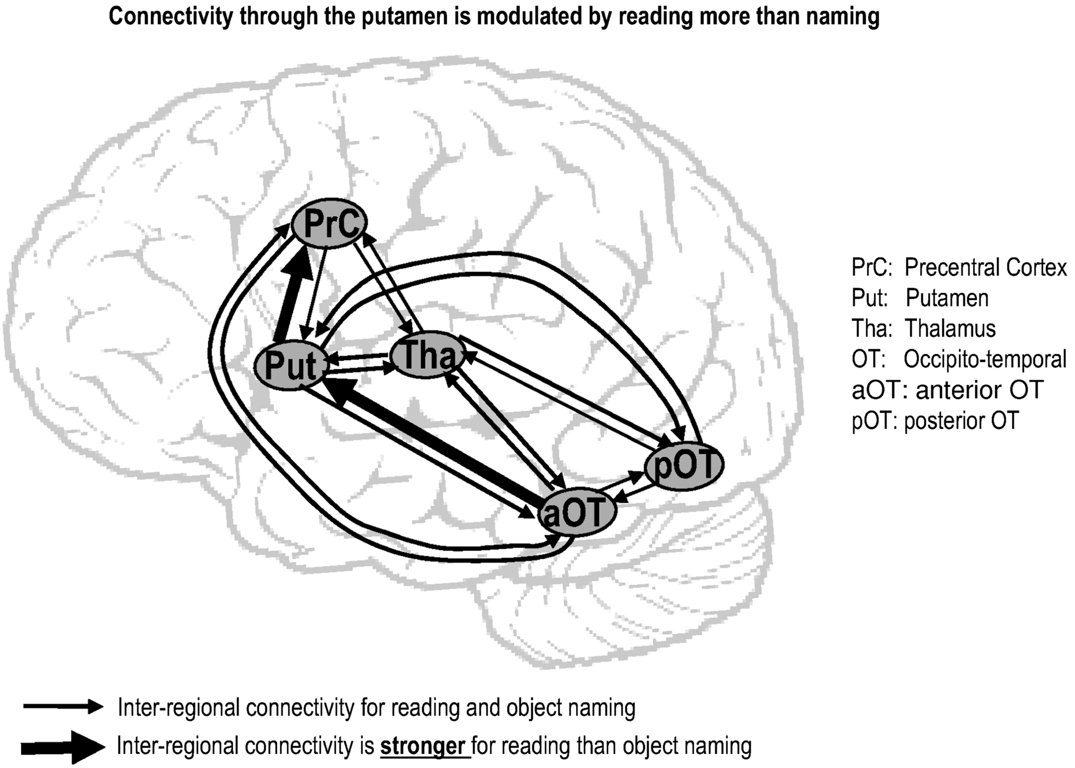
Connectivity through the putamen is modulated by reading more than naming
PrC
Put
Tha
pOT
aOT
PrC: Precentral Cortex
Put: Putamen
Tha: Thalamus
OT: Occipito-temporal
aOT: anterior OT
pOT: posterior OT
Inter-regional connectivity for reading and object naming
Inter-regional connectivity is stronger for reading than object naming

Figure 1. *An example of how dynamic causal modelling (DCM) can address modular and distributed architectures. This DCM includes five regions that are commonly activated during reading and picture naming. The results of the DCM analysis show that the connection from visual recognition areas (pOT/aOT) to articulatory areas (PrC), via the putamen, is stronger for reading than for object naming (Seghier & Price, 2010).*

neuroimaging (cf. McIntosh, 2004). We return to this later but first consider how the key regions (nodes) in casual modelling are identified in the first place.

Additive factors logic and context sensitivity

Prior to the inception of modern brain mapping, the principle of functional segregation was a hypothesis, based upon decades of careful electrophysiological and anatomical research (Zeki & Shipp, 1988). Within months of the introduction of whole-brain activation studies, the selective activation of functionally segregated brain areas was evident, and the hypothesis became a principle (e.g., Zeki et al., 1991). It is interesting to look back at how these activation maps were obtained experimentally: Most early brain mapping studies used an elaboration of Donder's subtractive method (e.g., Petersen, Fox, Posner, Mintun, & Raichle, 1988). Put simply, this entailed adding a process to a task and subtracting the evoked brain activity to reveal the activation due to the extra processing. Our first misconception was that one could associate the brain activation with the added process.

This interpretation of an activation rests upon the assumption of *pure insertion*—namely, that the extra process can be inserted purely without changing existing processes or eliciting new processes (or processing). The pure insertion assumption is very similar to the additivity assumption in the combination of factors in the additive factors method (Sternberg, 2011, this issue). In other words, it is necessary to exclude interactions between the old and new task factors before associating any brain activation with the new task component. To address this empirically, one needs a factorial design in which one can test for the interactions (Friston et al., 1996). The adoption of factorial designs—and the ability to assess interactions in neuroimaging experiments—was incredibly important and allowed people to examine the context-sensitive nature of the activations, in the sense of quantifying the dependency of the activation due to one factor on that of another. Factorial designs are now the mainstay of experimental design in neuroimaging. Certainly, in our unit, it has been nearly a decade since we have used a design that had fewer than two factors. Indeed, at the time of writing, a search on PubMed.gov for "interaction AND brain AND (fMRI OR PET)" yielded 1,854 results, while a search for "subtraction AND brain AND (fMRI OR PET)" gave only 1,615. Figure 2a shows a simple example of a factorial design and a test for a regionally specific interaction, again focusing on the processes underlying reading and object naming.

Cognitive subtraction and process decomposition

As described carefully by Sternberg (2011, this issue), there is a key distinction between cognitive subtraction and process decomposition. In cognitive subtraction, one changes the task in a qualitative way to induce a new putative processing component. Conversely, in process decomposition, the task remains the same but the stimuli or context is changed in a multifactorial way. This circumvents the assumption of pure insertion, while affording the opportunity to test for interactions. As noted in Sternberg (2011 this issue) "with a composite measure factorial experiments are essential, to assess how the effects of the factors combine". Neuronal responses are, by their nature, composite, in the sense that they reflect the processing of multiple processing elements. Figure 2b provides an example of a factorial design where stimulus factors were varied parametrically to reveal an interaction or dependency between name frequency and stimulus modality (pictures or written names). Much of the additive factors method rests upon excluding an interaction to make inferences about the decomposition of the underlying processes: If two processes do not interact, they can be decomposed functionally. Exactly the same logic underpins cognitive conjunctions in neuroimaging (Price & Friston, 1997). In cognitive conjunction analyses, one tests for colocalized activation attributable to two or more factors *in the absence of an interaction*. Figure 2 provides a simple example of this

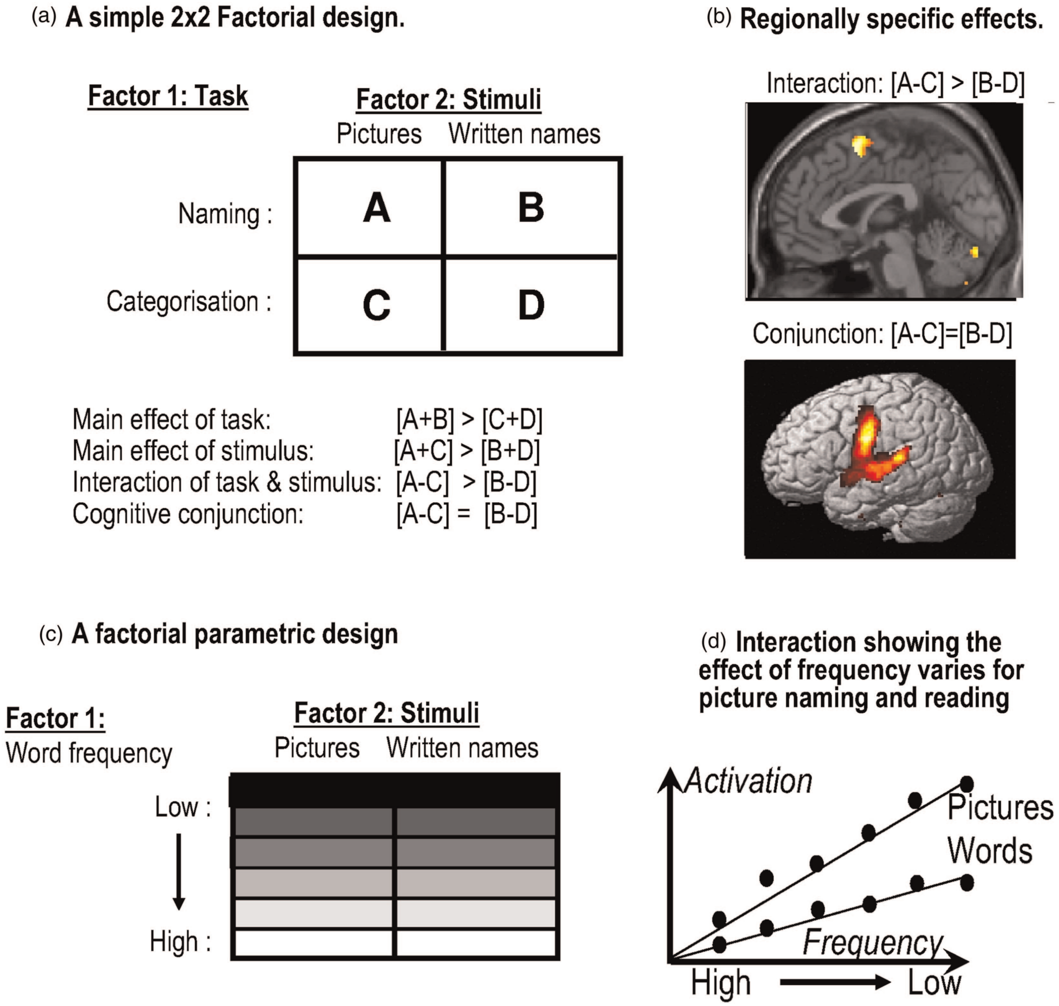
(a) A simple 2x2 Factorial design.
Factor 1: Task
Factor 2: Stimuli
Pictures
Written names
Naming :
Categorisation :
A
B
C
D
Main effect of task: [A+B] > [C+D]
Main effect of stimulus: [A+C] > [B+D]
Interaction of task & stimulus: [A-C] > [B-D]
Cognitive conjunction: [A-C] = [B-D]
(b) Regionally specific effects.
Interaction: [A-C] > [B-D]
Conjunction: [A-C]=[B-D]
(c) A factorial parametric design
Factor 1:
Word frequency
Factor 2: Stimuli
Pictures
Written names
Low :
High :
(d) Interaction showing the effect of frequency varies for picture naming and reading
Activation
Pictures
Words
Frequency
High
Low

Figure 2. *Factorial designs in brain activation experiments. (a) An example of a simple factorial design that uses the interaction to identify regions where differences between naming and semantic categorization are greater for pictures of objects than for their written names. (b) Regions identified by the interaction and conjunction (unpublished data). (c) An example of a factorial parametric design that uses the interaction to identify regions where the effect of a parametric factor (e.g., word frequency) is stronger for naming than for reading. (d) Hypothetical data illustrating an interaction. To view a colour version of this figure, please see the online issue of the Journal.*

method, which has become popular—with 195 PubMed.gov results for (cognitive AND conjunction) AND (fMRI OR PET OR neuroimaging). However, in the context of brain mapping, interactions can be extremely informative about neuronal processing and are usually used to infer the integration of inputs from two or more modules (brain regions). This can be essential in understanding the coupling among brain regions and the nature of hierarchical and recursive message passing among and within levels of sensory processing hierarchies. We pursue this theme in the context of changes in coupling below.

Context-sensitive coupling

In the same way that factorial designs disclose interactions in terms of regional processing, they can also inform the context-sensitive nature of coupling between brain areas. Interactions simply mean a difference in a difference (e.g., how a response to one factor depends upon the response to another). If we replace one (psychological) factor with the (physiological) activity in a seed or reference brain area, then the ensuing interaction becomes a psychophysiological interaction (PPI; Friston et al., 1997). Roughly speaking, this PPI reports a significant change in the (linear) influence of the seed region on any significant target region, with different levels of the psychological factor. Although a simple analysis, this has been exploited in a large number of neuroimaging studies to look at how coupling between brain areas can change with brain state or set—with 222 PubMed.gov results for (psychophysiological interaction OR PPI) AND (fMRI OR PET OR imaging). The notion that connectivity (and implicit modularity) is itself state and activity dependent is crucial for understanding the dynamic repertoire of real brain networks. Furthermore, it reiterates the importance of thinking about modular function in a context-sensitive way.

This becomes important in a pragmatic sense when one tries to understand the remote effect of brain lesions on brain activity and responses. This is usually referred to in terms of diaschisis (from Greek, meaning "shocked throughout"). A particular form of diaschisis can emerge when the remote effect of a lesion is itself context dependent—in other words, where there is an abnormality of evoked responses, due to a remote lesion that is revealed in, and only in, some specific tasks or brain states. This has been referred to as "dynamic diaschisis" (Price, Warburton, Moore, Frackowiak, & Friston, 2001) and underscores the subtleties in understanding highly context-dependent and nonlinear exchanges between modular brain regions. An example of dynamic diaschisis is shown in Figure 3. In the final section, we pursue the effect of brain lesions on evoked responses and look more closely at the notion of modularity and segregation, in the context of structure–function relationships.

Modularity, structure, and function

For many people, the goal of neuroimaging is to understand the functional architecture of the brain in relation to particular tasks or cognitive processing. This understanding entails knowledge of the mapping between the brain's structure and its function. An important complement to brain activation studies are lesion–deficit studies. Here brain imaging is used to define a regionally specific brain insult, and its implicit functional specialization is inferred from the associated behavioural deficit. Many people in neuroimaging have noted the importance of the complimentary contributions of functional and structural imaging. For example, identifying a regionally specific lesion, in the context of a behavioural deficit, suggests that this region was necessary for performance. Initially, it was hoped that a combination of lesion–deficit mapping and functional activation studies would identify necessary and sufficient brain regions for a particular task or process (Price, Mummery, Moore, Frackowiak, & Friston, 1999). However, this ambition quickly turned out to be misguided: This is because it overlooked the ubiquitous many-to-one (degenerate) mapping between structure and function in biological networks (Edelman & Gally, 2001). Put simply, this means that two or more areas could fulfil the same task requirements. Extending this notion to high-order combinatorics means that there may be no necessary brain area for any particular process and therefore no "necessary and sufficient brain area". This was a fundamental insight, which mandated a revision of (our) approaches to lesion–deficit mapping and activation studies (Price & Friston, 2002).

Crucially, the possibility of degenerate structure–function mappings brings us back to factorial designs and additive factors logic. One can see this simply by considering the difference between two structure–function relationships: In the first mapping, processing is distributed over two regions (nodes). This means that damage to the first or the second will produce a deficit. Now consider the degenerate case, where either

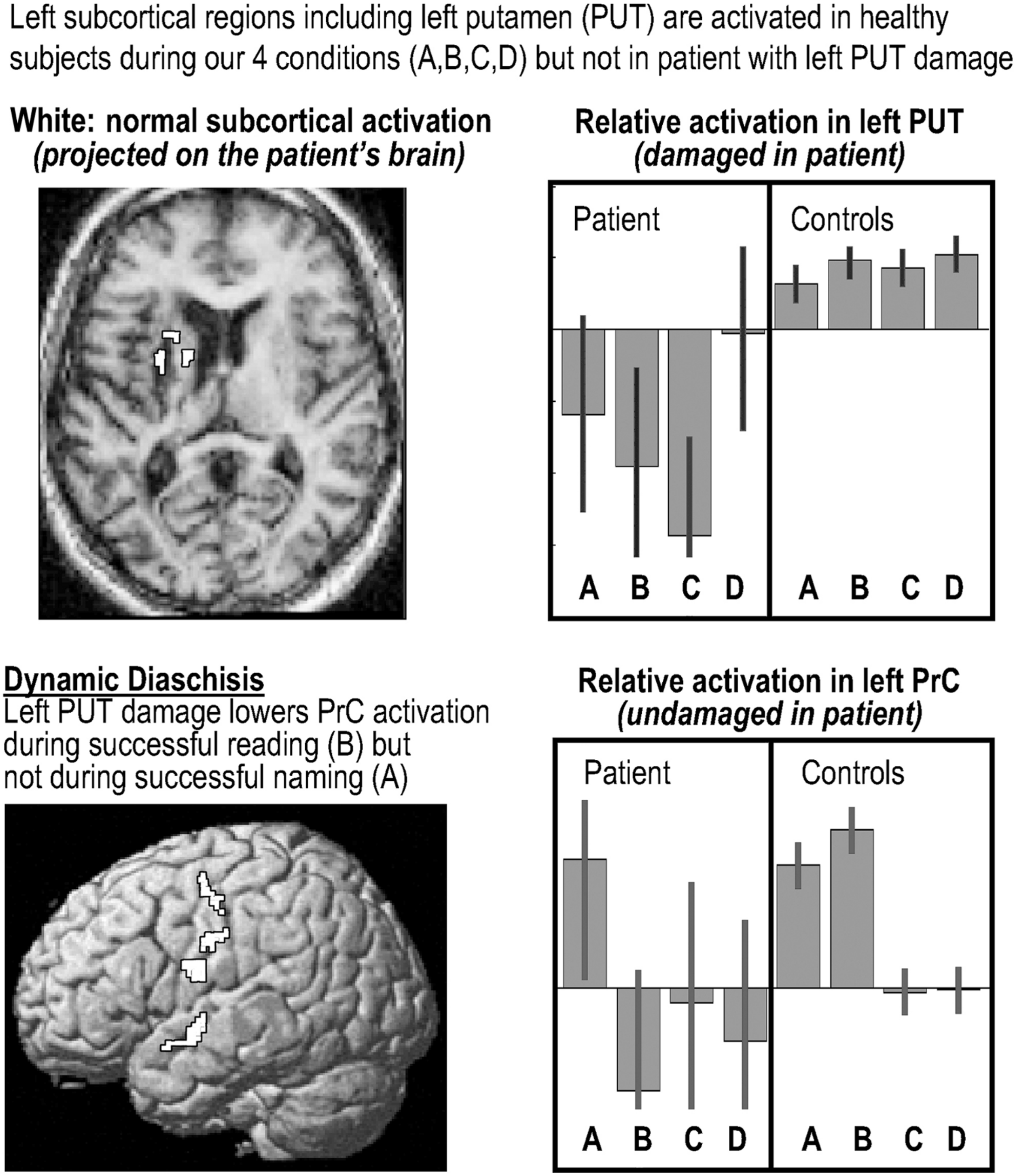

Figure 3. *An example of dynamic diaschisis (unpublished data). Following a lesion to the left putamen (see Figure 1), activation in the left precentral cortex (PrC in Figure 1) is abnormally low during successful reading but normally activated during successful object naming. This is consistent with the dynamic causal modelling (DCM) results reported in Figure 1 and suggests that PrC activation is driven by left putamen activation during reading but not naming. See Figure 2a for details of Conditions A, B, C, and D. To view a colour version of this figure, please see the online issue of the Journal.*

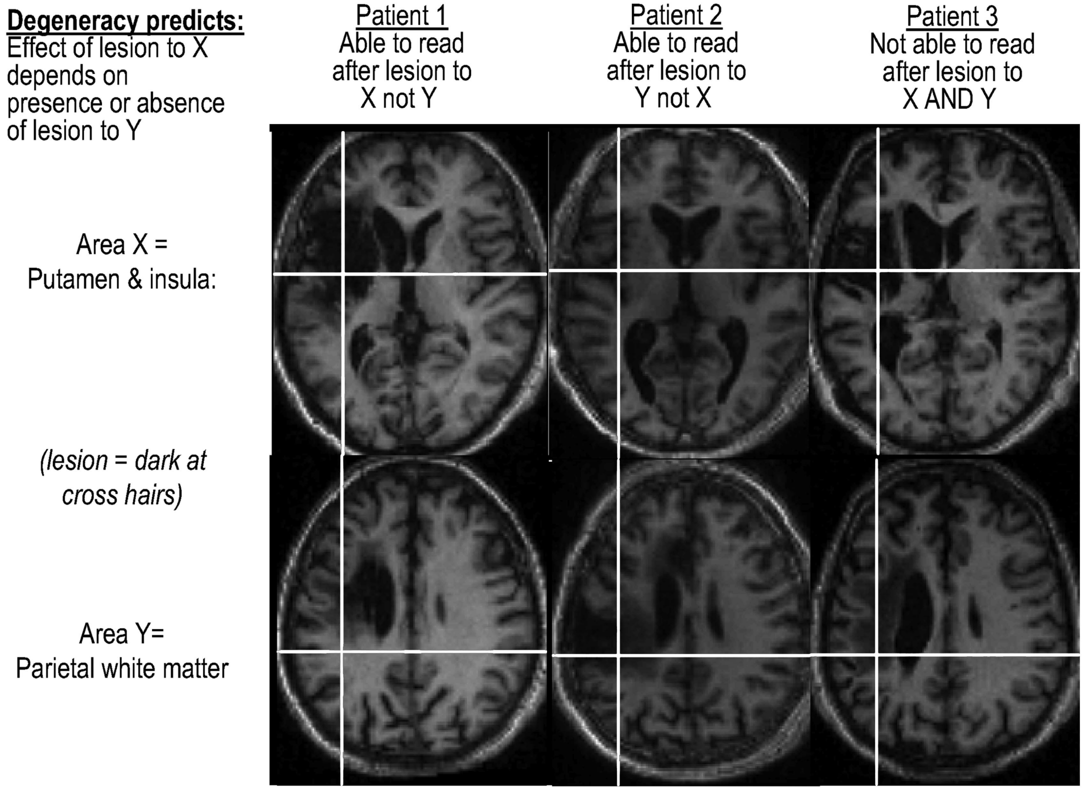
Degeneracy predicts:
Effect of lesion to X depends on presence or absence of lesion to Y
Area X = Putamen & insula:
(lesion = dark at cross hairs)
Area Y= Parietal white matter
Patient 1
Able to read after lesion to X not Y
Patient 2
Able to read after lesion to Y not X
Patient 3
Not able to read after lesion to X AND Y

Figure 4. *An example of degeneracy. Reading is impaired following lesions that damage the left putamen, left insula, and left parietal cortex inclusively (Patient 3). However, damage to only one of these regions does not impair reading (Patients 1 and 2). The results suggest that reading can be supported either by a pathway that involves the parietal cortex or by a pathway that involves the putamen/insula. When one pathway is damaged, the other pathway can support reading. When both pathways are damaged, reading is impaired. This previously unpublished result is consistent with a study of reading aloud in healthy subjects (Seghier, Lee, Schofield, Ellis, & Price, 2008) that showed an inverse relationship (across participants) between activation in the left putamen and parietal cortex. Together, the results from patient studies (above) and healthy subjects (Seghier et al., 2008) suggest that the putamen and parietal cortex are components of different reading pathways and that either one or the other is needed for successful reading.*

node can support the function. Here, only a lesion to the first *and* second area will cause a deficit. If we assume that the deficit is a pure measure of the assumed process in question, then we have two fundamentally different (multiplicative) combination rules within additive factors logic. In the first situation (deficit following lesions to first or second area), the probability of a deficit $p(D|L)$ is equal to one minus the probability that they are both undamaged, which is the product of the probability that neither are lesioned.

$$p(D|L)=1-[1-p(L_1=1)][1-p(L_2=1)] \qquad (1a)$$

Conversely, under the degenerate mapping, the probability of a deficit becomes the probability of a lesion in either area, leading to a very different combination rule:

$$p(D|L) = p(L_1 = 1)p(L_2 = 1) \qquad (1b)$$

Crucially, in order to disambiguate between these two scenarios we need a factorial design, in which we can lesion (or obtain access to patients with lesions in) one area $L_1 = 1$ and the other area $L_2 = 1$. In short, we need a multilesion analysis. This provides an important and principled motivation for studying patients with different brain lesions (and has implications for traditional single-case studies in neuropsychology). Figure 4 shows an example of degeneracy inferred using this Boolean logic associated with degenerate structure–function mappings. Interestingly, a classical one-to-one structure–function mapping implies that $p(D_i|L) = p(L_i = 1) \Rightarrow p(D_i|L) = p(D_i|L_i)$, which means the deficit is conditionally independent of lesion $L_k \in \{0, 1\} : \forall k \neq i$ in all other areas.

There are many interesting issues that attend the analysis of multilesions studies. However, we close by noting that a truly inclusive approach to modularity and structure–function mappings in the human brain will account for both lesion–deficit data and functional activation studies. In short, our empirical and conceptual models of brain architecture have to explain both evoked responses due to experimental manipulations in activation studies and the behavioural deficits elicited by selective lesions. Clearly, these models entail a precise specification of the mapping between neuronal activity and cognitive function. This mapping is itself a holy grail of cognitive neuroscience, which has been referred to as a cognitive ontology (Poldrack, 2006; Price & Friston, 2005). Indeed, cognitive ontologies are now becoming a major focus of the brain imaging and cognitive neuroscience community, particularly with the advent of new neuroinformatics tools (Poldrack, Halchenko, & Hanson, 2009). One can see how the combination of data from different modalities and different patients acquires a principled motivation from the arguments above.

In conclusion, the arguments and developments discussed in this review rest explicitly on the notion of modular but coupled brain regions and the additive factors method (with linear or nonlinear combination rules), introduced by Sternberg (2011 this issue).

REFERENCES

Bullmore, E., & Sporns, O. (2009). Complex brain networks: Graph theoretical analysis of structural and functional systems. *Nature Reviews. Neuroscience*, *10*(3), 186–198.

Calabretta, R. (2007). Genetic interference reduces the evolvability of modular and non-modular visual neural networks. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *362*(1479), 403–410.

Duffau, H. (2006). Brain plasticity: From pathophysiological mechanisms to therapeutic applications. *Journal of Clinical Neuroscience*, *13*(9), 885–897.

Edelman, G. M., & Gally, J. A. (2001). Degeneracy and complexity in biological systems. *Proceedings of the National Academy of Sciences, USA*, *98*(24), 13763–13768.

Friston, K. J., Büchel, C., Fink, G. R., Morris, J., Rolls, E., & Dolan, R. J. (1997). Psychophysiological and modulatory interactions in neuroimaging. *NeuroImage*, *6*(3), 218–229.

Friston, K. J., Frith, C. D., Liddle, P. F., & Frackowiak, R. S. (1993). Functional connectivity: The principal-component analysis of large (PET) data sets. *Journal of Cerebral Blood Flow and Metabolism*, *13*(1), 5–14.

Friston, K. J., Harrison, L., & Penny, W. (2003). Dynamic causal modelling. *NeuroImage*, *19*(4), 1273–1302.

Friston, K. J., Price, C. J., Fletcher, P., Moore, C., Frackowiak, R. S., & Dolan, R. J. (1996). The trouble with cognitive subtraction. *NeuroImage*, *4*(2), 97–104.

He, Y., & Evans, A. (2010). Graph theoretical modeling of brain connectivity. *Current Opinion in Neurology*, *23*(4), 341–350.

McIntosh, A. R. (2004). Contexts and catalysts: A resolution of the localization and integration of function in the brain. *Neuroinformatics*, *2*(2), 175–182.

Petersen, S. E., Fox, P. T., Posner, M. I., Mintun, M., & Raichle, M. E. (1988). Positron emission tomographic studies of the cortical anatomy of single-word processing. *Nature*, *331*(6157), 585–589.

Poldrack, R. A. (2006). Can cognitive processes be inferred from neuroimaging data? *Trends in Cognitive Sciences*, *10*(2), 59–63.

Poldrack, R. A., Halchenko, Y. O., & Hanson, S. J. (2009). Decoding the large-scale structure of brain function by classifying mental states across individuals. *Psychology Science*, *20*(11), 1364–1372.

Price, C. J., & Friston, K. J. (1997). Cognitive conjunction: A new approach to brain activation experiments. *NeuroImage*, *5*(4, Pt. 1), 261–270.

Price, C. J., & Friston, K. J. (2002). Degeneracy and cognitive anatomy. *Trends in Cognitive Sciences*, *6*(10), 416–421.

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

Price, C. J., Mummery, C. J., Moore, C. J., Frackowiak, R. S., & Friston, K. J. (1999). Delineating necessary and sufficient neural systems with functional imaging studies of neuropsychological patients. *Journal of Cognitive Neuroscience*, *11*(4), 371–382.

Price, C. J., Warburton, E. A., Moore, C. J., Frackowiak, R. S., & Friston, K. J. (2001). Dynamic diaschisis: Anatomically remote and context-sensitive human brain lesions. *Journal of Cognitive Neuroscience*, *13*(4), 419–429.

Redies, C., & Puelles, L. (2001). Modularity in vertebrate brain development and evolution. *Bioessays*, *23*(12), 1100–1111.

Seghier, M. L., Lee, H. L., Schofield, T., Ellis, C. L., & Price, C. J. (2008). Inter-subject variability in the use of two different neuronal networks for reading aloud familiar words. *NeuroImage*, *42*(3), 1226–1236.

Seghier, M. L., & Price, C. J. (2010). Reading boosts connectivity through the putamen. *Cerebral Cortex*, *20*(3), 570–582.

Sporns, O. (2010). *Networks of the brain*. Cambridge, MA: MIT Press.

Sternberg, S. (2011). Modular processes in mind and brain. *Cognitive Neuropsychology*, *28,* 156–208.

Valencia, M., Pastor, M. A., Fernández-Seara, M. A., Artieda, J., Martinerie, J., & Chavez, M. (2009). Complex modular structure of large-scale brain networks. *Chaos*, *19*(2), 023119. doi: 10.1063/1.3129783

Wagner, A. (2005). Robustness, evolvability and neutrality. *FEBS Letters*, *579,* 1772–1778.

Zeki, S., & Shipp, S. (1988). The functional logic of cortical connections. *Nature*, *335*(6188), 311–317.

Zeki, S., Watson, J. D., Lueck, C. J., Friston, K. J., Kennard, C., & Frackowiak, R. S. (1991). A direct demonstration of functional specialization in human visual cortex. *Journal of Neuroscience*, *11*(3), 641–649.