(forthcoming in *The British Journal for the Philosophy of Science*)

# Functional independence and cognitive architecture

**Vincent Bergeron**

## Abstract

In cognitive science, the concept of dissociation has been central to the functional individuation and decomposition of cognitive systems. Setting aside debates about the legitimacy of inferring the existence of dissociable systems from *behavioral* dissociation data, the main idea behind the dissociation approach is that two cognitive systems are dissociable, and therefore viewed as distinct, if each can be damaged, or impaired, without affecting the other system's functions. In this paper, I propose a notion of functional independence that does not require dissociability, and describe an approach to the functional decomposition and modelling of cognitive systems that complements the dissociation approach. I show that highly integrated cognitive and *neuro*cognitive systems can be decomposed into *non*-dissociable but functionally independent components, and argue that this approach can provide a general account of cognitive specialization in terms of stable structure-function relationship.

## 1 Introduction

In cognitive science, the concept of dissociation has been central to the functional individuation and decomposition of cognitive systems. The main idea behind the dissociation approach is that two cognitive systems are dissociable, and therefore viewed as distinct, if each can be damaged, or impaired, without affecting the other system's functions. More precisely, consider two cognitive systems *A* and *B*. If *B* can be damaged (or impaired) without affecting *A*'s functions, and *A* can be damaged (or impaired) without affecting *B*'s functions, then *A* and *B* are dissociable (and distinct) systems (even if the two might share some of their components).

In cognitive science, dissociable systems can be *inferred* on the basis of *behavioral* dissociations (or task dissociations). For example, a double dissociation between face recognition and visual object recognition—i.e. observing a patient with intact visual objects recognition but impaired face recognition, and another patient with intact face recognition but impaired visual object recognition—suggests that the system used to recognize faces is not identical to the system used to recognize objects, and that each of the two systems has at least one subsystem that the other doesn't have (Coltheart [2001]). In other words, this behavioral double dissociation suggests that the face recognition and visual object recognition systems are dissociable, and that each system can function independently from at least one subsystem of the other system. It would not, however, be reasonable to infer from this behavioral double dissociation that the face recognition and visual object recognition systems are *completely* distinct (or disjoint, see Lyons [2003]), since the two systems evidently share some of their subsystems (e.g. the subsystems responsible for low-level visual feature analysis).

This approach has a long history in neuropsychological research dating back to the early days of neurolinguistics when Paul Broca used it to infer the existence of a distinct speech articulation center in what is now known as Broca's area ([1861]). In the twentieth century, Hans-Lukas Teuber ([1955]) introduced the term 'double dissociation of function' into modern neuropsychology, and Tim Shallice later provided an extensive analysis of the dissociation logic (Shallice [1988]). Today, the (double) dissociation method remains one of the most powerful tools used to infer the existence of dissociable cognitive systems (Vallar [2000]; Coltheart [2001]), and it is widely used to infer the existence of cognitive *modules* since the latter are, in the weakest sense, dissociable (or separately modifiable) functional components (Carruthers [2006]; Sternberg [2011]; Shallice & Cooper [2011]).[1]

That being said, the legitimacy of inferring dissociable cognitive systems from behavioral double dissociation data is very much a matter of debate (Bergeron [2007]; Coltheart [2001]; Davies [2010]; Dunn & Kirsner [1988]; Juola, & Plunkett [2000]; Machery [2012]; Van Orden, Pennington, & Stone [2001]). Behavioral double dissociations are taken as evidence of dissociable systems, because cognitive

---

[1] Another popular approach to the functional individuation and decomposition of cognitive systems (or processes) is the *separate modifiability* approach (Sternberg [2011]). The main idea behind this approach is that two cognitive (sub)processes *A* and *B* can be viewed as distinct if each can be changed (or modified) independently of the other. Evidence for the separate modifiability of *A* and *B* is provided if one finds two factors *F* and *G* (e.g. experimental manipulations) that influence *A* and *B* selectively, that is, 'a change in the level of *F* influences *A* but leaves *B* invariant, while a change in the level *G* influences *B* but leaves *A* invariant' (p. 151). Both dissociability and separate modifiability have been used to provide minimal notions of cognitive modularity. For example, Carruthers ([2006]) states that 'in the weakest sense, a module can just be something like: a dissociable functional component' (p.2), and Sternberg ([2011]) states that 'two sub-processes *A* and *B* of a complex process (mental or neural) are modules if and only if each can be changed [modified] independently of the other' (p. 159).

architectures composed of dissociable systems would produce behavioral double dissociations if damaged in different ways (Coltheart [2001]).[2] This is an inference to the best explanation (as opposed to a deductive inference) because cognitive architectures composed of *non*-dissociable systems may also produce similar behavioral dissociations if damaged in different ways (see e.g. Juola & Plunkett [2000], Plaut [1995]).

Leaving aside the issue of what can be inferred from behavioral dissociation data, there is general agreement among philosophers and cognitive scientists on what needs to be explained. The large body of behavioral dissociation data indicates that there are specialized cognitive (sub)systems and that there are specific relations between these (sub)systems and particular cognitive functions. It is the identification of such distinct structure-function relationships that forms the basis for the functional decomposition and modelling of cognitive systems.[3]

In this paper, I propose a notion of functional independence that does not require dissociability (section 1), and describe an approach to the functional decomposition and modelling of cognitive systems that complements the dissociation approach (section 2). In particular, I show that highly integrated cognitive and *neuro*cognitive systems can be decomposed into *non*-dissociable but functionally independent components. I then argue

---

[2] Suppose that two cognitive systems $A$ and $B$ are dissociable because $A$ has a subsystem $S_a$ that $B$ doesn't have and $B$ has a subsystem $S_b$ that $A$ doesn't have. If $S_a$ is damaged while $B$ is left intact, then we should expect performance on behavioral task $T_a$ (which depends on $A$) to be impaired while performance on behavioral task $T_b$ (which depend on $B$) *not* to be impaired. Similarly, if $S_b$ is damaged and $A$ is left intact, then we should expect performance on behavioral task $T_b$ (which depends on $B$) to be impaired while performance on behavioral task $T_a$ (which depend on $A$) *not* to be impaired.

[3] More recently, functional neuroimaging data (e.g. fMRI, PET), in the form of selective activations of brain areas for certain tasks, also point to the wide range of specialized neural circuits in the brain, although the methodology in this case differs from the standard behavioral dissociation logic in neuropsychology. See Machery ([2012]) for a critical discussion of both methodologies.

that this approach can provide a general account of cognitive specialization in terms of stable structure-function relationship (section 3).

## 2. Functional Independence Without Dissociability

To see how a cognitive system can be functionally independent from another cognitive system without the two being dissociable, consider first the distinction between the low-level computational operations, or “workings”, performed by a cognitive system, and the higher-level cognitive “uses” to which it is put (Bergeron [2008]; Anderson [2010]). One can specify the cognitive workings of a cognitive system by identifying the kinds of computational operations that system is able to perform, and one can specify the cognitive uses of that system by identifying the cognitive functions that require the performance of any of these kinds of computational operations. The distinction is not between levels of analysis (e.g. cognitive, neural, etc.) but between different senses of the term ‘function’; it applies equally to both cognitive systems characterized functionally and *neuro*cognitive systems that are realized in particular brain areas.

To illustrate this idea, consider the case of Broca’s area (BA). On the one hand, we know that BA is put to a number of linguistic and non-linguistic *uses*—for example, it is involved in both musical and linguistic syntactic processing, in object manipulation, and in action sequencing and action perception (Maess *et al.* [2001], Nishitani et al. [2005]). On the other hand, there is mounting evidence that BA’s contribution to these various cognitive uses could be performed by a “reusable” set of low-level computational operations, or *workings*—e.g. sequencing operations on a wide range of inputs (Fiebach

& Schubotz [2006]), or the processing (detection, extraction) of hierarchical structures in a wide range of cognitive domains (Tettamanti & Weniger [2006]). It is clear from this example that BA's "function" can be interpreted in two different ways depending on whether one is referring to its local workings or to its cognitive uses. [4]

Consider now one of the cognitive uses of BA, namely, speech production. In order to contribute to this cognitive capacity, BA needs to access and compute over information that is processed and made available by other brain areas, one of which is the superior temporal sulcus (STS) which processes and stores phonological representations (Hickok & Poeppel [2007]). Therefore, relative to speech production, BA and the STS are *not* dissociable, since BA's contribution to speech production—e.g. sequencing operations on phonological representations—would be affected if the STS were damaged.[5] Nevertheless, there is a sense in which BA is functionally independent from the STS.

BA is functionally independent from the STS *with respect to (e.g.) sequencing operations* in the sense that BA performs these operations and could do so even if the STS did not perform anything. Generalizing:

> (FI): A system *S* is *functionally independent* from another system *O with respect to working W*, iff *S* performs *W* and could perform *W* even if no part of *O* performed any workings.[6]

---

[4] I discuss the functional anatomy of Broca's area in more detail in section 3.

[5] Notice that the STS's contribution to speech production would also be affected if BA were damaged.

[6] This formulation is inspired from the analysis of isolability provided by Lyons ([2001]): 'A substrate *S* is *isolable with respect to task T* iff *S* performs task *T* and could do so even if nothing else computed any (cognitive) functions'(p. 289 original emphasis.) (FI) and Lyons' concept of isolability differ in that (FI) is a two-place relation between cognitive

Accordingly:

> (FI systems) Two systems *S* and *O* are *functionally independent systems* with respect to workings $W_s$ and $W_o$, iff *S* and *O* are functionally independent of each other.

To say, in (FI), that *S could perform W* even if no part of *O* computed anything is to say that *S* possesses the right kind of machinery (i.e. internal structure, mechanism, etc.) to perform *W*, in the sense that *if* it were to have access to the right kind of information, it would perform *W*. The second condition for functional independence must be read counterfactually.

In the case in which *S* performs *W* over information provided by *O*, this means that *S* has the right kind of machinery to perform *W* over the kind of information that is normally made available by *O*. For example, if we assume that BA does in fact perform sequencing operations over phonological representations that are processed and made available by the STS, then to say that BA could perform these operations even if no part of the STS computed anything is to say that BA would do so if we could find some other way of supplying it with the kind of information that the STS normally provides.

---

systems while isolability is a one-place property of cognitive systems. Also, (FI) does *not* provide a *general* way of individuating cognitive systems (more on this in section 2 below). By contrast, Lyons' notion of isolability is part of a general account of cognitive systems that is perfectly compatible with both the functional independence and dissociation approaches—some pairs of Lyonsian systems are functionally independent of each other (but not dissociable from each other), some are dissociable (but not functionally independent), and some are both functionally independent and dissociable (see Lyons [2001].)

This idea, that a cognitive system *S* which performs working *W* on information made available by another system *O* could perform *W* even if *O* failed to compute anything, is also consistent with (and helps us to make sense of) cases in which a working of *S* is put to the service of multiple cognitive functions. Consider once again the case of BA. As mentioned earlier, evidence suggests that the computational work done by BA may be put to a number of linguistic and non-linguistic uses. Thus, even if one of the systems that BA needs to access is damaged, such that BA can no longer be put to one of its cognitive uses (e.g. damage to the STS prevents BA's contribution to speech production), BA may continue to perform the very same kinds of computations in the service of other cognitive functions (e.g. BA could still contribute to action sequencing and object manipulation by accessing and performing sequencing operations on motor representations held in various motor areas).[7]

Finally, to say, in (FI systems), that two systems *S* and *O* are functionally independent of each other with respect to workings $W_s$ and $W_o$ is *not* to say that *S* and *O* are *specialized for* $W_s$ and $W_o$, respectively. In other words, it is not to say that $W_s$ and $W_o$ are the *only* cognitive workings that *S* and *O* respectively perform. I discuss the notion of a specialized system in section 3 below.

## 3. FI Systems and Cognitive Architecture

One of the main challenges for the study of cognitive architecture has been to give a notion of functional components that can support functional decomposition without

---

[7] This point is considered again in section 3 below.

losing sight of cognitive integration. Ideally, such a notion should be sufficiently weak (or have enough flexibility) to support the functional decomposition of a wide range of cognitive systems. In particular, it should apply to the functional analysis of a cognitive system, whether or not the components of that system are dissociable. That being said, a notion of functional components must also be sufficiently strong to explain the phenomena that have traditionally motivated the study of cognitive architecture. As noted earlier, the large number of behavioral dissociations in the neuropsychological literature—and more recently the large number of functional neuroimaging studies showing the selective activation of brain regions for particular cognitive tasks—points to the existence of a wide range of specialized cognitive systems, and the functional decomposition and modelling of these systems requires that we be able to identify the *distinct* structure-function relationship that exists for each of their functional components. Finally, a notion of functional components should (ideally) be general enough to support the functional decomposition and modelling of cognitive systems, whether these are characterized purely functionally or mapped onto the brain. The notion of FI systems appears to be a good candidate for the task.

To begin with, there is a sense in which FI is weaker than the notion of dissociability. As we saw in the previous section, relative to a particular cognitive task, two cognitive systems (or subsystems) can be functionally independent (in FI's sense) without being dissociable. Accordingly, FI is sufficiently weak to support the functional decomposition and modelling of cognitive systems that have *non*-dissociable functional components.

This point is particularly important when considering interactive activation models of cognition, a class of computational models in which there is continuous and bidirectional transmission of information between functional components (McClelland & Rumelhart [1981]). Consider, for example, the dual-route cascaded (DRC) model of visual word recognition and reading aloud (Coltheart *et al.* [2001]). The model, which is depicted in figure 1, consists of three routes (two lexical routes and one non-lexical route), and each route is composed of a number of interacting components. On the lexical nonsemantic route, for instance, the visual features of a word's letters activate the word's letter units, which then activate the word's entry in the orthographic input lexicon. This word entry then activates the corresponding word entry in the phonological output lexicon, which then activates the word's phonemes. What is particular about the process is that with the exception of the visual feature units, the processing of information in each of the route's components is continuously influenced, via bidirectional excitatory and inhibitory connections, by the processing occurring in at least one other component of the model.

INSERT FIG. 1 HERE

For example, the DRC model takes into account the classic finding that the perception and recognition of a letter by human subjects can be facilitated by presenting it in the context of a word (Reicher [1969]). In the model, this phenomenon is accounted for by the presence of both excitatory and inhibitory feedback connections between the letter units and orthographic input lexicon components. For instance, a string of printed letters activates the visual feature units component, which then begins the process of activating the letters, in the letter units component, that are consistent with the visual inputs. These partially activated letters then begin to activate the words, in the orthographic input lexicon component, that are consistent with the letters. These activated words then produce excitatory feedback, in the letter units component, that reinforces the activation of the letters in the activated word(s), and inhibitory feedback that suppresses the activation of the letters that are not in the word(s). This kind of interaction continues until a lexical decision is made.

As this example shows, the letter units and orthographic input lexicon components of the DRC model are not dissociable, since we would expect the functioning of the letter units component to be impaired if the orthographic input lexicon component were damaged, a prediction that has been empirically confirmed (Coltheart *et al.* [2001]). Yet, the two components are functionally independent of each other (in FI’s sense), as each possesses the right internal structure and algorithms, and could thus continue to perform its computational workings even if the other component failed, provided that we found another way to supply it with the relevant information. Since this is also true of any pair of components on the lexical nonsemantic route, these can be characterized as FI (sub)systems.

There is a sense, however, in which FI is stronger than the notion of dissociability. As noted earlier, two cognitive systems that share some of their subsystems can dissociate, if each system contains at least one subsystem that the other doesn’t have, and damage occurs only to these subsystems that are not shared. For example, recall that a double dissociation between face recognition and object recognition tasks suggests that the face recognition and object recognition systems are dissociable even though the two systems evidently share some of their components. Two such systems would not, however, be functionally independent in FI’s sense, since according to FI, two systems cannot both share subsystems and be functionally independent. Suppose that *A* and *B* are two cognitive systems that share some of their subsystems. *A* would not be functionally independent from *B* since FI requires that *A* could perform its workings even if *no part* of *B* computed anything. But if no part of *B* computed anything, *A* could not perform its workings, since some of its parts (the subsystems it shares with *B*) would not be computing anything.

The fact that the notion of functional independence proposed here precludes the sharing of parts between functional entities is the reason why it is sufficiently strong to support the functional decomposition and modelling of highly integrated cognitive systems. As previously noted, this kind of functional analysis requires that we be able to identify the *distinct* structure-function relationship that exists for each functional component of a cognitive system. What the DRC model example illustrates is that this can be achieved if the components of the model are characterized as FI subsystems, or to put it differently, if each component can be characterized as a distinct structure-*working* relationship.

We have just seen that (FI) is, in some sense, both weaker and stronger than the notion of dissociability. It is weaker in that (FI) does not require dissociability, and it is stronger in that two cognitive systems can be dissociable without being functionally independent (in FI's sense). The two notions are therefore quite distinct, and neither one implies the other. Now we can see precisely how the functional independence approach complements the dissociation approach.

As we noted earlier, the notion of dissociability provides an intuitive way of individuating cognitive systems. If two cognitive processes *A* and *B* can be impaired independently of each other (i.e. if the two are dissociable), then it makes sense to hypothesize that *A* and *B* rely on different systems, even if *A* and *B* appear to share some of their subprocesses.[8] But as we also saw, the various functional components of a given cognitive system may not be dissociable, in which case we need a different way of functionally individuating these components. The notion of FI systems provides this additional tool for the functional decomposition and modelling of cognitive systems.[9]

Moving on now to our last desideratum for a notion of cognitive components, the notion of FI systems should apply equally well to the functional decomposition and modelling of *neuro*cognitive systems, in which the functional components are mapped onto particular brain structures. To illustrate how the notion of FI systems can support this kind of analysis, consider the development of the two-level model of working memory operations in the lateral prefrontal cortex (Petrides [1996], [2005]).

Working memory is broadly defined as the capacity for online maintenance and manipulation of stored information during a wide range of cognitive activities (Baddeley [1986]). In monkeys, working memory processes have long been known to depend on the normal functioning of the frontal lobes, as evidenced by several early lesion studies. In particular, research in the 1950s, 60s and 70s has shown that lesions to the lateral portion of the prefrontal cortex (PFC), and not to various other cortical areas, render monkeys incapable of performing otherwise simple delayed-response tasks. In humans, several subsequent neuropsychological studies carried out in the 1970s and 1980s, and involving tasks similar to the ones used with monkeys, have confirmed that working memory

---

[8] Keeping in mind, as noted earlier, that this inference is at best an inference to the best explanation.

[9] Below I describe how the functional independence approach works with particular methods in cognitive neuroscience (e.g. lesion studies, functional neuroimaging.)

processes similarly depend on the normal functioning of the lateral PFC (see Fuster [2008] for a review of both monkey and human studies).

In the mid-1990s, Michael Petrides proposed a model of the contribution of the lateral PFC to working memory processes (Petrides [1996]). According to the model, different areas of the lateral PFC make distinct contributions to these cognitive functions. The mid-ventrolateral PFC, in interaction with posterior cortical association areas where long-term storage and transient maintenance of information occurs, mediates basic executive functions such as the active retrieval, selection, and comparison of stimuli held in short- and long-term memory. In contrast, the mid-dorsolateral PFC, in interaction with the mid-ventrolateral PFC, mediates more complex executive functions such as the monitoring and manipulation of information within working memory.[10] Let us refer to these two levels of executive functions as the cognitive *uses* of these two areas.

The model then explains the fundamental difference in the contribution of the mid-dorsolateral and mid-ventrolateral PFC to working memory processes by specifying two distinct sets of local computational operations (cognitive workings). The first one, which occurs in the mid-ventrolateral PFC, consists of the initiation of active retrieval and active encoding of information held in short- and long-term memory. These operations are crucial for any cognitive task that requires the active retrieval, selection, and comparison of mnemonic information. The second one, which occurs in the mid-

---

[10] A classic example of these more complex executive functions is the self-ordered task. In this task, subjects are presented, on each trial, with different arrangements of the same set of stimuli (e.g. abstract visual patterns, pictures of objects, words), and are required to select a different stimulus on each trial until all the stimuli have been selected. Thus, successful performance of this task requires that the subjects keep track of which stimuli have, or have not, been selected from trial to trial, a capacity that requires the constant comparison of previously selected stimuli with non-selected ones.

dorsolateral PFC, consists of the re-coding in abstract form, and online maintenance, of multiple pieces of information held in short- and long-term memory. These operations are crucial for any cognitive task that requires the monitoring and manipulation of information within working memory.[11]

It is clear from the model that the cognitive workings of both the mid-dorsolateral and mid-ventrolateral PFC are not dissociable from the cognitive workings of the other cortical areas with which they interact—the mid-ventrolateral PFC can't retrieve or encode any mnemonic information that is not processed and made available by the posterior cortical association areas, and the mid-dorsolateral PFC can't re-code or maintain online any mnemonic information if the mid-ventrolateral PFC can't retrieve and encode it in the first place. Yet, both areas are functionally independent (in FI's sense) from the other areas with which they interact—the mid-ventrolateral PFC maintains the *capacity* to initiate the retrieval and encode mnemonic information regardless of what happens to posterior cortical areas, and the mid-dorsolateral PFC maintains the *capacity* to re-code mnemonic information and keep it online regardless of what happens to the mid-ventrolateral PFC. So these areas can be said to realize distinct FI (sub)systems.

---

[11] Experimental evidence for this model has come from several different lines of research in both macaque monkeys and humans. In monkeys, evidence comes from lesion studies and from electrophysiological recordings of the activity of single neurons in either mid-dorsolateral or mid-ventrolateral PFC during various working memory tasks. In humans, evidence comes from neuropsychological studies on patients with selective damage to parts of the lateral PFC, and from functional neuroimaging studies in which the experimental tasks can be specifically designed to recruit the executive control operations that the model attributes to the two mid-lateral prefrontal cortical regions. See Petrides ([2005]) for a review of the evidence.

Notice also that the different cognitive workings that the model ascribes to the mid-ventrolateral and mid-dorsolateral PFC are specified domain-*neutrally*. This is because these cognitive workings remain the same regardless of cognitive use. In fact, the model makes two predictions in the form of two domain-neutral structure-working relationships. First, any cognitive task that requires the initiation of active retrieval and active encoding of mnemonic information should recruit the mid-ventrolateral PFC. Second, any cognitive task that requires the re-coding in abstract form and online maintenance of multiple pieces of mnemonic information should recruit the mid-dorsolateral PFC. These two predictions have been extensively tested, and confirmed, in humans using either positron emission tomography (PET) or functional magnetic resonance imaging (fMRI). For example, the first prediction was tested in studies with normal subjects performing verbal (Petrides *et al.* [1995]), spatial (Owen *et al.* [1996]), and non-spatial (Petrides *et al.* [2002]) working memory tasks, all of which required a certain degree of active retrieval processing. Performance on these tasks, relative to a matching control task (no active retrieval requirement) resulted in significantly greater activity within the mid-ventrolateral PFC. Similarly, the second prediction was tested in studies with normal subjects performing a non-spatial visual self-ordered task (Petrides *et al.* [1993a]), verbal self-ordered and externally ordered tasks (Petrides *et al.* [1993b]), and spatial and non-spatial *n*-back tasks (Braver *et al.* [1997]; Owen *et al.* [1998]). Performance on these tasks (with monitoring requirements) relative to a visual matching control task (with no monitoring requirement) resulted in significantly greater activity within the mid-dorsolateral PFC.

The testing of these two predictions nicely illustrates how experimental evidence for the existence of particular FI subsystems can be obtained using functional neuroimaging methods. Here, the cognitive contribution of individual brain areas (specified as structure-working relationships or FI subsystems) can be tested in a relatively straightforward way using imaging data. There are, of course, important methodological issues that arise with such functional neuroimaging experiments (Klein [2010]; Machery [2012]), but the point here is simply that the functional independence approach comports naturally with existing methods in cognitive neuroscience.

To sum up then, the proposed notion of functional independence offers a conception of cognitive components (as FI subsystems) that can support the functional decomposition and modelling of a wide range of cognitive systems, whether we approach the study of these systems at the purely functional level or in terms of their underlying neural architectures.

## 4. FI Systems and Cognitive Specialization

Our discussions of both the DRC model of visual word recognition and reading aloud and the two-level model of working memory are instructive because they show that highly integrated cognitive systems can be decomposed into non-dissociable but functionally independent components, or FI (sub)systems. Another important, and related, issue for the study of cognitive architecture is whether the fact that some of the components of such systems appear to have multiple cognitive uses in different domains—e.g. Broca's area and the mid-ventrolateral and mid-dorsolateral PFC—precludes the existence of cognitive specialization. That is, whether this fact precludes the identification of *stable* structure-working relationships. I now argue that the existence of cognitive (or neurocognitive) specialization is perfectly compatible with the existence of such “multifunctional” cognitive components.[12]

One might indeed expect, in the course of evolution or normal development, that the workings of some cognitive components established for one purpose may remain fixed while being put to new uses (and often without losing their original uses). Sometimes this could be due to substantial changes in the overall system’s organization, but sometimes this could simply result from the establishment of new functional connections between preexisting components.

There is, in fact, mounting evidence that a great many brain structures are recruited by different tasks across different cognitive domains (Anderson [2010]; Poldrack [2006]), which suggests that the “reuse” of neural circuitry for various cognitive purposes could be a central organizational principle of the brain[13].

Anderson ([2010]) recently observed that the apparent multifunctionality of many brain areas poses a significant challenge to the characterization of neurocognitive specialization.

---

[12] As my discussions most naturally apply to *neuro*cognitive systems, I will focus on the notion of neurocognitive specialization, although the precise notion of cognitive specialization that will emerge applies to both the neural and purely functional levels of analysis.

[13] Anderson ([2010]) provides an integrated review of recent neural reuse theories. In particular, see Vittorio Gallese’s ‘neural exploitation hypothesis’ ([2008]), Susan Hurley’s ‘shared circuits model’ ([2008]), Stanislas Dehaene's ‘neuronal recycling’ theory ([2005]) and Michael Anderson's ‘massive redeployment’ hypothesis ([2007]).

> if selectivity and localization are not in fact central features of the functional organization of the brain, how shall we think about the function-structure relationship? (p. 246)

The characterization of multifunctional brain areas as FI subsystems may help provide a solution to this problem. Instead of characterizing the specialization of a brain area on the basis of its (domain specific) cognitive uses—e.g. characterizing the fusiform face area as a face recognition module (Kanwisher *et al.* [1997])—we should attempt to characterize it domain-neutrally by specifying its reusable cognitive workings, e.g. via multiple- or cross-domain investigations.

It has, however, proven difficult to provide precise characterizations of the local computational operations performed by individual brain areas. Some (e.g. Coltheart [2013]) even argue that the fact that cognitive neuroscientists are not yet able to provide such precise characterizations effectively prevents functional neuroimaging data from informing cognitive theories—how can the activation or non-activation of a brain area be used to support (or argue against) a particular cognitive theory, one might ask, if we don't know what kind of computations the area specializes in? In this respect, it is particularly helpful to note another advantage of characterizing the components of neurocognitive systems as FI subsystems.

The computational properties of particular brain structures are more easily studied in animal models. In vision research, for example, a great deal of our current knowledge of human visual areas comes from an extensive mapping and neurophysiological investigation of the macaque monkey's visual systems (Van Essen *et al.* [1992]).

Similarly, as the development of the two-level model of working memory demonstrates, significant advances in our understanding of the mechanisms underlying working memory functions come from the neurophysiological and neuropsychological investigation of the macaque monkey's lateral prefrontal cortex. In other words, the human brain shares many of its functions with that of other species, and for any human cognitive function, we can expect that (at least) some component(s) of it could be found in the cognitive repertoire of another species. What is less clear, however, is how best to exploit this evolutionary continuity in building models of human cognition.

As we have seen, neural reuse theories suggest that in the course of evolution, some brain structures may acquire a number of cognitive uses while maintaining their cognitive workings fixed. This, in turn, suggests that *homologous* brain structures may contribute to very different cognitive capacities and hence have very different cognitive uses, while basically sharing the same low-level internal operations or workings. This is, essentially, homology thinking applied to brain function (Bergeron [2010]).

The idea of functional homology may seem confused at first (Love [2007]). After all, the concept of homology was originally defined as "the same organ in different animals under every variety of form and function" (Owen & Cooper [1843], p. 379), where sameness is defined by common phylogenetic origin. However, as just noted, the fact that homologous brain structures appear functionally dissimilar based on a comparison of their cognitive uses obscures the fact that they may share the same workings. Thus, by specifying the cognitive workings of homologous structures domain-neutrally, as neural reuse theories suggest we do, one could test whether this is in fact the case.[14]

In light of this, let us define the notion of *cognitive homology* as the cognitive workings of homologous brain structures serving (potentially) different cognitive uses in different animals. The search for cognitive homologies could then greatly benefit the task of characterizing the cognitive workings of individual human brain areas, since the cognitive workings of homologous brain structures are likely to be similar (when characterized domain-neutrally) because of their shared ancestry.

For example, the recent discovery of the homologue of Broca's area in the macaque monkey (Petrides *et al.* [2005]), combined with the fact that the two areas share the same architectonic and neurophysiological properties, suggests that a more precise characterization of Broca's area's cognitive workings could be achieved in the macaque monkey. Similarly, several studies in the past fifteen years indicate that the basic cytoarchitecture of the lateral prefrontal cortex in the human and macaque monkey brains, as well as the anatomical connectivity of its various architectonic areas, is comparable (Petrides *et al.* [2012]). This, in turn, suggests that a more precise characterization of the cognitive workings that the two-level model ascribes to the mid-dorsolateral and mid-ventrolateral PFC could be achieved in the macaque monkey's homologues of these areas.[15]

---

[14] See also Moore ([2013]) for a discussion of the concept of homology applied to developmental psychology.

[15] For example, studies in monkeys suggest that the online maintenance of multiple pieces of mnemonic information could be realized by the presence of reverberatory neural circuits in the mid-dorsolateral PFC (Tegnér *et al.* [2002]).

The identification of specialized neurocognitive (sub)systems (*stable* structure-working relationships), however, requires more than the attribution of low-level computational operations to particular brain areas. One must also ascertain that the computational operations that are attributed to an area (whatever these might be) be the *only* types of operations it performs. More precisely:

> (SS) A system *S is specialized for* working(s) *W*, iff *S* performs *W*, or part of *W*, whenever it (or part of it) is cognitively active.[16]

(SS) captures the idea that a system may be specialized for more than one *closely related* working. For example, in the above-described two-level model of working memory, the mid-dorsolateral PFC is specialized for the abstract re-coding *and* online maintenance of pieces of information, which would imply that whenever it (or part of it) is active, this area is performing at least one of these two types of operations. Similarly, we can say that the lexical nonsemantic route of the DRC model is specialized for visual word recognition *and* phonological conversion, which is to say that this system (or part of it) is performing at least part of these operations whenever it is active (although there might not be any brain area, or network of brain areas, that is specialized for these operations). Ultimately, whether it makes sense to say that a system is specialized for multiple cognitive workings will depend on the particular theoretical context.

---

[16] Equivalently: 'A substrate *S* specializes in task *T* iff *T* is an exhaustive specification of the cognitive input/output function that *S* computes' (Lyons [2001], p. 289). If *W* is an exhaustive specification of the cognitive input/output function that S computes, then *S* must be performing *W*, or part of *W*, whenever it (or part of it) is cognitively active. And, if *S* performs *W*, or part of *W*, whenever it (or part of it) is cognitively active, then *W* must be an exhaustive specification of the cognitive input/output that *S* computes.

Now part of the problem with the identification of specialized neurocognitive (sub)systems is that brain areas can be individualized in several different ways. The three most common ways are 1- on the basis of cytoarchitectonic analysis (e.g. Brodmann's areas); 2- on the basis of gross anatomical features (e.g. gyri, sulci, and nuclei); and 3- on the basis of functional analysis (e.g. using neurophysiological, neuropsychological, and functional neuroimaging data). Consider, once again, the case of Broca's area. The classical definition of this area is represented in Brodmann's cytoarchitectonic map as areas 44 and 45. However, recent anatomical and electrophysiological studies point to several regional differences in connectivity, architectonics, and functional properties of Broca's area (Amunts & Zilles [2012]). One may therefore hypothesize that (at least some of the) distinct subregions within the classically defined Broca's area support different *types* of computational operations. If this were true, it would probably make more theoretical sense to say that different subregions of Broca's area are each specialized for a different cognitive working than to say that Broca's area is specialized for several different types of cognitive workings.

A second hypothesis (not incompatible with the first) is that some of the anatomically distinct subregions within Broca's area perform the *same* type of computations in a recursive fashion. Support for this hypothesis comes from the recent demonstration that the distinct subregions of Broca's area are organized in a hierarchical fashion (Amunts *et al.* [2010]), combined with the fact (mentioned earlier) that the cognitive tasks to which Broca's area is put (e.g. speech production, action sequencing, planning behavior) have been conceptualized as hierarchically organized.[17] Under this hypothesis, like in the first, stable structure-working relationships could be found for each of the distinct subregions of Broca's area, although in this case multiple subregions would be specialized for the same computational operations.[18]

The "multifunctionality" of brain areas, therefore, does not necessarily preclude the identification of specialized neurocognitive (sub)systems (or stable structure-working relationships). The functional independence approach is well suited for the specification of the reusable workings of highly integrated neurocognitive components, a task that is made easier by studying the detailed anatomical and functional properties of the homologues of human brain areas in other species.

Vincent Bergeron
Department of philosophy
University of Ottawa
55 Laurier Ave. East, room 8101
Ottawa ON Canada, K1N 6N5
vincent.bergeron@uottawa.ca

## Acknowledgements

Thanks to Max Coltheart, Dustin Stokes, and two anonymous referees for helpful comments on earlier drafts of this paper. Parts of this paper were presented at the

---

[17] Another possibility, less consistent with neural reuse, is that different subregions within Broca's area perform the same type of computations simultaneously (in parallel) on different, but unique, input domains. In this case, the neurocognitive specialization of each subregion/neural population could be characterized domain-specifically. I thank an anonymous referee for pointing out this possibility.

[18] There is nothing surprising about the idea that multiple brain structures could share the same neurocognitive specialization (at least when characterized domain-neutrally). For example, if we assume that at least some of the basic principles of auditory and visual analysis are similar, then we should expect to find parts of the auditory cortex that perform the same (or very similar) workings as parts of the visual cortex.

Annual Meeting of the European Society for Philosophy and Psychology in 2012, and at the Agent Tracking and Its Disorders Conference at Macquarie University in 2011. I am grateful to members of both audiences for comments and criticisms.

## References

Amunts, K., Lenzen, M., Friederici, A. D., Schleicher, A., Morosan, P., Palomero-Gallagher, N., and Zilles, K. [2010]: 'Broca's region: novel organizational principles and multiple receptor mapping', *PLoS Biology* **8**, e1000489.

Amunts, K. and Zilles, K. [2012]: 'Architecture and organizational principles of Broca's region', *Trends in Cognitive Sciences*, **16**, pp. 418–26.

Anderson, M.L. [2007]: 'The massive redeployment hypothesis and the functional topography of the brain', *Philosophical Psychology*, **21**, pp. 143–74.

Anderson, M. [2010]: 'Neural reuse: A fundamental organizational principle of the brain', *Behavioural and Brain Sciences*: **33**, pp. 245–66.

Bergeron, V. [2007]: 'Anatomical and Functional Modularity in Cognitive Science: Shifting the Focus', *Philosophical Psychology*, **20**, pp. 175–95.

Bergeron, V. [2008]: *Cognitive architecture and the brain: Beyond domain-specific functional specification*, doctoral dissertation, Department of Philosophy, University of British Columbia, <http://circle.ubc.ca/handle/2429/2711>

Bergeron, V. [2010]: 'Neural Reuse and Cognitive Homology', *Behavioral and Brain Sciences*, **33**, pp. 268–9.

Braver, T. S., Cohen, J. D., Nystrom, L. E., Jonides, J., Smith, E. E., and Noll, D. C. [1997]: A parametric study of prefrontal cortex involvement in human working memory*, Neuroimage*, **5**, pp. 49–62.

Broca, P. [1861]: 'Remarques sur le siège de la faculté du langage articulé, suivis d'une observation d'aphémie (perte de la parole)' *Bulletin de la Société Anatomique*, **6**, pp. 330–57.

Carruthers, P. [2006]: *The Architecture of the Mind*, Oxford University Press.

Coltheart, M. [2001]: 'Assumptions and methods in cognitive neuropsychology', In B. Rapp (*ed.*), *The handbook of cognitive neuropsychology*, Psychology Press, pp. 3–

21.

Coltheart, M. [2013]: ‘How Can Functional Neuroimaging Inform Cognitive Theories?’, *Perspectives on Psychological Science*, **8**, pp. 98–103.

Coltheart, M., Rastle, K., Perry, C., Langdon, R., and Ziegler, J. [2001]: ‘DRC: a dual route cascaded model of visual word recognition and reading aloud’, *Psychological Review*, **108**, pp. 204–56.

Davies, M. [2010]: ‘Double Dissociation: Understanding Its Role in Cognitive Neuropsychology’, *Mind and Language*, **25**, pp. 500–40.

Dehaene, S. [2005]: ‘Evolution of human cortical circuits for reading and arithmetic: The ‘neuronal recycling’ hypothesis’, In S. Dehaene, J. Duhamel, M.D. Hauser and G. Rizzolatti (*eds*), *From Monkey Brain to Human Brain*, MIT Press.

Dunn, J. C. & Kirsner, K. [2003]: ‘What can we infer from double dissociations?’, *Cortex*, **39**, pp. 1–7.

Fiebach, C. J. and Schubotz, R. I. [2006]: ‘Dynamic anticipatory processing of hierarchical sequential events: a common role for Broca's area and ventral premotor cortex across domains?’, *Cortex*, **42**, pp. 499–502.

Fuster, J. M. [2008]: *The prefrontal cortex*, Academic Press.

Gallese, V. [2008]: ‘Mirror neurons and the social nature of language: The neural exploitation hypothesis’, *Social Neuroscience*, **3**, pp. 317–33.

Hickok, G. and Poeppel, D. [2007]: ‘The cortical organization of speech processing’, *Nature Reviews Neuroscience*, **8**, pp. 393–402.

Hurley, S. [2008]: ‘The shared circuits model (SCM): How control, mirroring, and simulation can enable imitation, deliberation, and mindreading’, *Behavioral and Brain Sciences*, **31**, pp. 1–58.

Juola, P. and Plunkett, K. [2000]: ‘Why double dissociations don't mean much’, In G. Cohen, R. A. Johnston, and K. Plunkett (*eds*), *Exploring cognition: damaged brains and neural networks*, Psychology Press, pp. 319–27.

Kanwisher, N., McDermott, J., and Chun, M. M. [1997]: ‘The fusiform face area: a module in human extrastriate cortex specialized for face perception*, Journal of Neuroscience*, **17**, pp. 4302–11.

Love, A. C. [2007]: ‘Functional homology and homology of function: Biological concepts and philosophical consequences' *Biology and Philosophy*, **22**, pp. 691–708.

Lyons, J. C. [2001]: 'Carving the Mind at its (Not Necessarily Modular) Joints', *British Journal for the Philosophy of Science*, **52**, pp. 277–302.

Lyons, J. C. [2003]: 'Lesion studies, spared performances, and cognitive systems', *Cortex*, **39**, pp.145–7.

Machery, E. [2012]: 'Dissociations in neuropsychology and cognitive neuroscience', *Philosophy of Science*, **79**, pp. 490–518.

Maess, B., Koelsch, S., Gunter, T. C., and Friederici, A. D. [2001]: 'Musical syntax is processed in Broca's area: an MEG study', *Nature Neuroscience*, **4**, pp. 540–5.

McClelland, J. L., and Rumelhart, D. E., [1981]: 'An interactive activation model of context effects in letter perception: part 1. An account of basic findings', *Psychological Review*, **88**, pp. 375–407.

Moore, D. S. [2013]: 'Importing the Homology Concept from Biology into Developmental Psychology', *Developmental Psychobiology*, **55**, pp. 13–21.

Nishitani, N., Schürmann, M., Amunts K. and Hari, R. [2005]: 'Broca's region: From action to language', *Physiology*, **20**, pp. 60–9.

Owen, R. & Cooper, W. W. [1843]: 'Lectures on the comparative anatomy and physiology of the invertebrate animals, delivered at the Royal College of Surgeons, in 1843, Longman, Brown, Green, and Longmans.

Owen, A. M., Evans, A. C., and Petrides, M. [1996]: 'Evidence for a two-stage model of spatial working memory processing within the lateral frontal cortex: a positron emission tomography study', *Cerebral Cortex*, **6**, pp. 31–8.

Owen, A. M., Stern, C. E., Look, R. B., Tracey, I., Rosen, B. R., and Petrides, M. [1998]: 'Functional organization of spatial and nonspatial working memory processing within the human lateral frontal cortex', *Proceedings of the National Academy of Science USA*, **95**, pp. 7721–26.

Petrides, M. [1996]: 'Specialized systems for the processing of mnemonic information within the primate frontal cortex', *Philosophical Transactions of the Royal Society of London B*, **351**, pp. 1455–61.

Petrides, M. [2005]: ‘Lateral prefrontal cortex: architectonic and functional organization’ *Philosophical Transactions of the Royal Society of London B*, **360**, pp. 781–95.

Petrides, M., Alivisatos, B., and Evans, A. C. [1995]: ‘Functional activation of the human ventrolateral frontal cortex during mnemonic retrieval of verbal information’, *Proceedings of the National Academy of Science USA*, **92**, pp. 5803–07.

Petrides, M., Alivisatos, B., Evans, A. C.,and Meyer, E. [1993a]: ‘Dissociation of Human Mid-Dorsolateral from Posterior Dorsolateral Frontal-Cortex in Memory Processing’, *Proceedings of the National Academy of Sciences USA*, **90**, pp. 873–7.

Petrides, M., Alivisatos, B., and Frey, S. [2002]: ‘Differential activation of the human orbital, mid-ventrolateral, and mid-dorsolateral prefrontal cortex during the processing of visual stimuli’, *Proceedings of the National Academy of Sciences USA*, **99**, pp. 5649–54.

Petrides, M., Alivisatos, B., Meyer, E., and Evans, A. C. [1993b]: ‘Functional activation of the human frontal cortex during the performance of verbal working memory tasks’, *Proceedings of the National Academy of Sciences USA*, **90**, pp. 878–82.

Petrides, M., Cadoret, G. V., and Mackey, S. [2005]: ‘Orofacial somatomotor responses in the macaque monkey homologue of Broca's area’, *Nature*, **435**, pp. 1235–38.

Petrides, M., Tomaiuolo, F., Yeterian, E. H., and Pandya, D. N. [2012]: ‘The prefrontal cortex: comparative architectonic organization in the human and macaque monkey brains’, *Cortex*, **48**, pp. 46–57.

Plaut, D. C. [1995]: ‘Double dissociation without modularity: evidence from connectionist neuropsychology’, *Journal of Clinical & Experimental Neuropsychology*, **17**, pp. 291–321.

Poldrack, R. A. [2006]: ‘Can cognitive processes be inferred from neuroimaging data?’, *Trends in Cognitive Science*, **10**, pp. 59–63.

Reicher, G. M. [1969]: ‘Perceptual recognition as a function of meaningfulness of stimulus material’, *Journal of Experimental Psychology*, **81**, pp. 274–80.

Shallice, T. [1988]: *From Neuropsychology to Mental Structure*, Cambridge University Press.

Shallice, T. and Cooper, R. P. [2011]: *The organization of mind*, Oxford University Press.

Sternberg, S. [2011]: ‘Modular processes in mind and brain. Cognitive Neuropsychology’, **28**, pp. 156–208.

Tegnér, J., Compte, A., and Wang, X-J [2002]: ‘The dynamical stability of reverberatory neural circuits’, *Biological Cybernetics*, **87**, pp. 471–81.

Tettamanti, M. and Weniger, D. [2006]: ‘Broca's area: a supramodal hierarchical processor?’, *Cortex*, **42**, pp. 491–4.

Teuber, H. L. [1955]: ‘Physiological psychology’, *Annual Review of Psychology*, **9**, pp. 267–96.

Vallar, G. [2000]: ‘The methodological foundations of human neuropsychology: studies in brain-damaged patients’, In F. Boller and J. Grafman (*eds*), *Handbook of Neuropsychology*, Elsevier, pp. 305–44.

Van Essen, D. C., Anderson, C. H., and Felleman, D. J. [1992]: ‘Information processing in the primate visual system: an integrated systems perspective’, *Science*, **255**, pp. 419–23.

Van Orden, G. C., Pennington, B. F., and Stone, G. O. [2001]: ‘What do double dissociation prove?’, *Cognitive Science*, **25**, pp. 111–72.

FIGURE CAPTION

Fig. 1. Coltheart et al. (2001) dual route cascaded model of visual word recognition and reading aloud (reproduced with permission form the American Psychological Association).