ABSTRACT

My dissertation applies philosophical analysis to the problem of how we should cognitively characterize brain activity. Let us distinguish between high-level cognitive functions—e.g. decision-making, face recognition—and the lower-level computational operations that are carried out by discrete regions of the brain. One can assume that cognitive functions are assembled from interactions between relatively autonomous computational operations carried out by discrete brain regions. My thesis, stated very broadly, is that in order to be effective, the decomposition of a cognitive function into a set of interactions between localized computational operations may need to be specified domain-neutrally, and not in terms of a particular informational domain or stimulus class.

Jerry Fodor’s influential work on modularity has sparked an industry of research that is based on the idea that the mind is, to a large extent, a configuration of domain-specific and relatively autonomous cognitive mechanisms, or modules. My treatment indicates how this modular approach must be modified in order successfully to decompose domain-specific cognitive functions into localizable computational operations. I proceed in two steps. First, I provide an analysis of the kinds of inferences that are used by cognitive scientists to postulate the existence of cognitive modules; I call these the modularity inferences. I offer a new characterization of these inferences, and argue that they can, and do, operate in three distinct modes in cognitive scientific research. Second, I present a general approach to the decomposition of a cognitive function into localizable computational operations. According to this
approach, which I call the working zone approach, the contribution of a distinct brain region to a cognitive function is specified in terms of the type of operations that this region performs, and not in terms of a particular informational domain. I demonstrate the value of this approach in several research contexts within the cognitive sciences.
# TABLE OF CONTENTS

Abstract ........................................................................................................... ii

Table of contents ......................................................................................... iv

List of figures ................................................................................................. vi

Acknowledgements ....................................................................................... vii

Dedication ....................................................................................................... viii

Part One: The modular approach to the study of cognition ................. 1

1 Anatomical and functional modularity in cognitive science ............... 2
   I. The modular approach to the study of cognition ............................... 2
   II. Anatomical and functional modularity ............................................ 14
   III. Case study 1: the face-recognition module hypothesis ............... 21

2 On imputing functions to cognitive structures .................................. 27
   I. Working vs. use: two distinct modes of functional specification .... 28
   II. A new perspective on the modularity inferences ......................... 33
   III. Case study 2: Wernicke’s insight ............................................... 35

3 The anatomical and functional modularity inferences: three 
   operating modes ............................................................................... 49
   I. Three operating modes .............................................................. 49
   II. An ambiguity in the notion of cognitive module ....................... 58

Part Two: The working zone approach and integration in cognitive 
neuroscience ............................................................................................... 61

4 The working zone approach ................................................................. 62
   I. Cognitive working zones ............................................................... 64
   II. A taxonomy of functional specifications for cognitive neuroscience .. 67
   III. Case study 3: the lateral prefrontal cortex and working memory .... 71
5 Integration in cognitive neuroscience ....................................................... 83
   I. Horizontal and vertical integration .................................................. 85
   II. Functional neuroimaging and cognitive theory ............................. 89
   III. Evolutionary considerations and animal models of human cognition .... 100
   IV. Conclusion ..................................................................................... 107

6 Etiological versus non-etiological functional specification ....................109
   I. A further ambiguity in the meaning of ‘function’ ............................. 109
   II. Etiological versus non-etiological functional specification ............... 115

References .............................................................................................. 129
LIST OF FIGURES

Figure 1  Coltheart et al. (2001) dual route cascaded model of visual word recognition and reading aloud .................................................. 30

Figure 2  A taxonomy of functional specifications for cognitive neuroscience: four distinct kinds of form-‘function’ relationships ................. 68

Figure 3  Etiological versus non-etiological workings and uses. Four distinct notions of function .......................................................... 113
ACKNOWLEDGEMENTS

I would like to thank Mohan Matthen for his constant and generous support over the past several years. I have benefited greatly from his knowledge and advice while working on this dissertation. Mohan’s energy and marked enthusiasm for my project have been truly inspiring, and at times crucial. I would also like to thank Dominic McIver Lopes for his sustained presence and encouragements over the course of my degree. I owe a great deal of what I have accomplished as a graduate student at UBC to my numerous and fruitful interactions with Dom. Looking back, I can safely say that on many occasions where I have made the right decision while working on this project it is because of Mohan and Dom’s sound advice.

Finally, I am grateful to Max Coltheart, Christopher Stephens, Dustin Stokes, and Patrick Rysiew for their helpful comments and suggestions on parts of this dissertation.
À Madeleine
PART ONE

The Modular Approach to the Study of Cognition
I. The modular approach to the study of cognition

Much of contemporary cognitive science is committed to what we might call the modular approach to the study of cognition. Jerry Fodor (1983), who was very influential in establishing the concept of a module, says: “One day . . . Merrill Garrett made what seems to me to be the deepest remark that I have yet heard about the psychological mechanisms that mediate the perception of speech. ‘What you have to remember about parsing is that basically it’s a reflex.'” (Dedication). Reflexes are quick, inflexible, and involuntary responses to stimuli, and Fodorian modules are cognitive reflexes. In its most recent and general form, the core of this approach involves viewing the architecture of human cognition, or at least part of it, as a configuration of quick specialized cognitive mechanisms, or subsystems, that are largely functionally independent of one another, that are localised in definite regions of the brain, and that operate over a distinct domain of cognitive information.

The earliest and most famous supporter of this view was Franz Joseph Gall (1758-1828). Gall himself did not use the words ‘module’ and ‘cognitive information’; however, his account of the nature and organization of human mental faculties nicely fits the above description of the core idea behind the
modular approach. That being said, we must carefully distinguish Gall’s view of the mind as a collection of independent ‘faculties’, from his preferred method of investigation, namely phrenology. While his view of the mind consists of the assumption that it is composed of functionally independent mechanisms that operate over a restricted domain of information—let us call any such assumption about the general architecture of the mind a **modularity assumption**—his method of investigation consists of a particular inferential strategy that Gall thought could be used successfully to elucidate the specific nature of such a modular architecture, *if the human mind/brain were to be thus organized*—let us call such inferential strategies the **modularity inferences**. In what follows, I will use the term ‘modular approach’ to refer to any research strategy (or framework) which consists of postulating cognitive modules on the basis of a particular modularity assumption and by making use of a particular kind of modularity inferences.

Recent discussions of cognitive modularity have been shaped in some important ways by Jerry Fodor’s *The Modularity Of Mind* (Fodor, 1983). In this book, which is by far the most cited work on the subject to appear since Gall, Fodor had two main goals. The first was to offer a rigorous analysis of the notion of cognitive module, and the second one was to propose a new general model of human cognitive architecture. Let us start with the former.

The first thing that must be noted concerning Fodor’s account of cognitive modularity is that contrary to a common portrayal of his project, Fodor did *not* provide a definition (in terms of necessary and sufficient conditions) of what a
cognitive module is, nor did he attempt to do so (see Coltheart, 1999 for a detailed discussion of this point). What he did, instead, was to suggest a number of system properties that are likely to be associated with modular systems:

I am not, in any strict sense, in the business of defining my terms... So what I propose to do instead of defining ‘modular’ is to associate the notion with a pattern of answers to such questions as... 1- Is [the system] domain specific, or do its operation cross content domains?... 2- Is the computational system innately specified, or is its structure formed by some sort of learning process? 3- Is the computational system ‘assembled’ (in the sense of having been put together from some stock of more elementary subprocesses) or does its virtual architecture map relatively directly onto its neural implementation? 4- Is it hardwired (in the sense of being associated with specific, localized, and elaborately structured neural systems) or is it implemented by relatively equipotential neural mechanisms? 5- Is it computationally autonomous, or does it share horizontal resources (of memory, attention, or whatever) with other cognitive systems? (Fodor, 1983, pp. 36-37)

Fodorian modules are typically domain specific, innately specified, not assembled, hardwired, computationally autonomous, informationally encapsulated (modules have access to only a very restricted amount of
information contained in the system as a whole), fast, cognitively impenetrable, they likely have a characteristic pattern of development, and their operation is likely mandatory. It is noteworthy how much of this characterization follows on Fodor’s reflex metaphor. Domain-specificity parallels the singularity of the stimulus that sets off a reflex; autonomy, mandatoriness, hardwiring, and encapsulation mirror the standard reflex-arc model. Fodor in fact explicitly states that none of these features should be regarded as a necessary requirement on modularity. ‘The notion of modularity ought to admit of degrees’ (p.37), and ‘the present contention, in any event, is relatively modest: it’s that if a psychological system has most of the modularity properties, then it is very likely to have all of them’ (p.137).

Fodor’s second goal was to propose a general account of human cognitive architecture. He argued for a bipartite organization of the mind into input analyzers (or input systems) and central systems, in which the input systems (roughly language and perceptual systems) are modular in the above sense-. The central systems (responsible for memory, attention, belief formation, and thought in general), for their part, are non-modular, and are best described as a collection of equipotential systems, i.e. brain systems in which each part can be trained or conditioned to perform a function previously (or actually) performed by another part of the system. Fodor’s case for the modularity of input systems has generally been well received by cognitive scientists and philosophers alike. This is not to say, however, that his sophisticated characterization of cognitive modules found its way unaltered into the actual practice (see below). His thesis
of the non-modularity (or equipotentiality) of central systems, however, quickly came under attack\(^1\).

There is no doubt that *The Modularity of Mind* has provided some fundamental insights into the architecture of human cognition. Both Fodor’s analysis of modularity as well as his bipartite picture of the mind have generated many fruitful discussions among cognitive scientists and philosophers alike. Moreover, his taxonomy of mental processes has helped renew the long standing debate over the best way to characterize the overall architecture of human cognition. One might want to ask, however, the extent to which these discussions have actually affected post-fodorian uses of the notion of cognitive module by cognitive scientists. That is, one might want to know how different (if different at all) the current working definition of cognitive modularity is from the conception which was prevalent in the early days of neuropsychology, or, for that matter, from the pre-fodorian days of twentieth century cognitive science.

A good indication that the working definition of cognitive modularity hasn’t changed much is that the main methodological tools for inferring cognitive modules haven’t changed. On the whole, cognitive modules are being hypothesized today using the same kind of logic that was used in 19\(^{th}\) century neuropsychology. True, the kinds of data that are used in making those inferences has expanded greatly as a result of the introduction of new investigative techniques—e.g. the new functional imaging and electrical

\(^1\) For two early critiques of Fodor’s thesis of the non-modularity of central systems see Marshall (1984) and Shallice (1984). For a more recent criticism, see Carruthers (2006). See also the discussion below.
recording techniques—but the overall logic of inferring cognitive modularity on
the basis of behavioral or behaviorally derived data is still the same (more on this
in section II and Chapter 2, section II).

A more direct way of assessing the impact of Fodor’s conception of modularity
in various branches of cognitive science is to get a sense of some of the leading
theorists’ reactions to his proposal. A review of the literature indicates that among
cognitive scientists, two main conceptions of cognitive modularity have emerged,
and that both of them can be seen as weakened versions of Fodorian modularity.

The first of these conceptions centers around the notions of functional and
neurological distinctness, and it is endorsed explicitly by Tim Shallice in his
influential book *From Neuropsychology to Mental Structure* (Shallice, 1988). The
basic argument behind this account comes from Herbert Simon (1969), and it
has been clearly stated by Marr (1976; 1982):

*Principle of modular design:* any large computation should be split up
and implemented as a collection of small sub-parts that are as nearly
independent of one another as the overall task allows. If a process is
not designed in this way, a small change in one place will have
consequences in many other places. This means that the process as a
whole becomes extremely difficult to debug or to improve, whether by
a human designer or in the course of natural evolution, because a
small change to improve one part has to be accompanied by many
simultaneous compensating changes elsewhere (Marr, 1976, p.485)
Although Simon and Marr’s computational (theoretical) argument is not specific to human psychology, many cognitive scientists consider it to be one of the two main justifications for adopting the modular approach. The other one is empirical, and it comes from neuropsychological research. Here, the presence of highly selective cognitive impairments (dissociations) like prosopagnosia (selective lost of the ability to recognise familiar faces) and various linguistic processing deficits suggest the functional independence of at least some cognitive processes.

Moreover, the fact that these dissociations often arise in patients showing fairly localized brain lesions seems to provide considerable evidence for the neurological specificity of these processes (more on this in sections II and III).

The most striking feature of the Marr-Shallice account of modularity is that there is no mention of domain specificity. Modules could, of course, be domain specific, but domain specificity is not seen as a requirement on modularity. Another distinctive feature of this notion of cognitive modularity is that the notion of functional independence (or distinctness) is only loosely defined, at least as compared to Fodor’s corresponding notions of computational autonomy, information encapsulation and cognitive impenetrability. A system is considered functionally independent from another system if the amount of interaction (or processing) between the two systems is relatively low compared to the amount of interaction (or processing) between the systems’ internal components, such that the integrity of one system can be compromised without necessarily compromising the integrity of the other.
Some readers may now wonder if this account of cognitive modularity does any theoretical work for cognitive scientists (see e.g. Elman et al., 1996). Indeed, it is so far from Fodor’s original proposal, and it is different enough from the most common usage of the term in cognitive science (see below), that one seems to be justified in asking whether it makes sense to call it an account of cognitive modularity.

The principal motivation for weakening the Fodorian account of modularity in this particular way has been a common rejection of Fodor’s input/central dichotomy (Marshall, 1984; Schwartz & Schwartz, 1984; Shallice, 1984; Shallice, 1988). One of The Modularity of Mind’s most controversial theses states that the non-modular nature of central processes (processes related to thought and the fixation of belief) is likely to be the reason why ‘cognitive science has in fact made approximately no progress in studying these processes’ (p.38). Towards the end of the book, Fodor, reflecting on the table of contents of the September 1979 issue of Scientific American which was devoted to the brain, sums up his argument:

[T]here is nothing on the neuropsychology of thought—presumably because nothing is known about the neuropsychology of thought. I am suggesting that there is a good reason why noting is known about it—namely, that there is nothing to know about it… [T]here are no content-specific central processes for the performance of which correspondingly specific neural structures have been identified.
Everything we know is compatible with the claim that central problem-solving is subserved by equipotential neural mechanisms (Fodor, 1983, p.119)

Fodor’s reasoning is that in so far as neuropsychology is the discipline concerned with linking particular mental operations to distinct areas of the brain, and that it has so far failed to do so in the case of central processes, it is reasonable to conclude that perhaps these processes are implemented by equipotential mechanisms, i.e. by multipurpose areas of the brain that are each capable of doing what every other area does. It should therefore be not too surprising if neuropsychologists have a hard time localizing particular central processes in specific areas of the brain.

Twenty five years later, it is fair to say that cognitive scientists who are primarily interested in understanding the structure of central processes have been generally unimpressed by Fodor’s remark. The study of central processes (attention, executive functions, memory functions) was, at the time The Modularity of Mind was published, and still is an integral part of neuropsychological research, and it now also flourishes within the cognitive neurosciences as well (see e.g. Andrewes, 2001; Eichenbaum, 2002; Stuss, & Knight, 2002). It is not that we have come to view these processes as content-specific, nor that we have come to realize that a great many of them are in fact content-specific, but rather, this fact simply reflects the widely held view among cognitive scientists that the relationship between brain structure and central
processes can be studied independently of the issue of content- or domain-
specificity ( Chapters 4 and 5 explore this approach in great details).

A second weakened version of Fodorian cognitive modularity has been
proposed by Max Coltheart (1999). According to this ‘neo-Fodorian’ account, a
module is

a cognitive system whose application is domain specific; here domain
specificity is a necessary condition for the applicability of the term
‘modular’ (p.118, my emphasis)

Coltheart then immediately follows with:

Now it is necessary to say something about what might be meant by
‘domain-specific’. I mean that a cognitive system is domain-specific if it
only responds to stimuli of a particular class: thus, to say that there is a
domain-specific face-recognition module is to say that there is a
cognitive system that responds when its input is a face, but does not
respond when its input is, say, a written word, or a visually presented
object, or someone’s voice (p.118)

The main difference between this proposed account of cognitive modularity and
the Marr-Shallice account is that domain-specificity is seen here as a necessary
condition for modularity\(^2\). This feature being the only necessary requirement for modularity, Coltheart argues, it is an empirical question as to whether the particular module proposed possesses any one of the other features of modularity described by Fodor. That being said, systems (or mechanisms) that are domain-specific will, almost by definition, show some level of functional independence, comparable at least to the level of functional independence which is characteristic of the Marr-Shallice account of modularity\(^3\).

Coltheart's conception of cognitive modularity can thus be seen as the Marr-Shallice conception plus domain-specificity. Moreover, it is fair to say that these two conceptions of cognitive modularity, taken together, account for most cases of the use of the terms 'module' and 'modular' in cognitive scientific research today. True, when they hypothesize the existence of cognitive modules, cognitive scientists will often complement their 'definition' of modularity with additional Fodorian features such as informational encapsulation and innate specificity—and they will also almost always make a reference to Fodor's book—but very few will actually provide empirical evidence for these additional features, over and above the evidence they provide for functional and neurological distinctness and/or domain-specificity.

Coltheart's conception of cognitive modularity, however, is by far the most popular of the two, at least when it comes to using the term 'module' (as opposed to the term 'component' or 'subsystem' for instance). As we will see throughout

\(^2\) In this respect, Coltheart's account of cognitive modularity may not be weaker, but stronger than Fodor's, whose account does not contain any necessary requirement for modularity.

\(^3\) Coltheart (1999) in fact argues that the operation of domain-specific cognitive systems is likely to be informationally encapsulated (to a certain degree), fast, mandatory, and hardwired.
this dissertation, it has been widely embraced (although rarely explicitly) in most areas of cognitive science, from the nineteenth century origin of modern neuropsychology to contemporary cognitive neuropsychology and cognitive neuroscience\(^4\). It is also the predominant conception of cognitive modularity in several branches of contemporary cognitive psychology, including developmental, and evolutionary psychology (see e.g. Barkow, Cosmides, & Tooby, 1992; Cosmides, & Tooby, 1994; Hirschfeld, & Gelman, 1994; Karmiloff-Smith, 1992; Pinker, 1994; Pinker, 1997; Sperber, 1994). For example, it is common among evolutionary psychologists to endorse some version of what Dan Sperber (1994) has called the *massive modularity thesis*. In its most extreme version, this thesis states that the mind is *entirely* composed of cognitive modules, where module is defined as a domain-specific cognitive mechanism (with the additional characteristic of being a distinct evolutionary adaptation). Gall’s picture of the mind was perhaps the earliest version of this view. In its more moderate (and popular) version, the massive modularity thesis states that the mind is *largely* composed of such cognitive modules. The latter have been postulated to account for cognitive capacities as diverse as theory of mind, face recognition, reading, cheating detection, grammar induction, and a variety of social understanding abilities. For these reasons, I will henceforth use the terms ‘module’ and ‘modular’ in Coltheart’s (1999) sense.

\(^4\) Neuropsychology is the study of the relationship between brain and behaviour. Both neuropsychology and cognitive neuropsychology thrive on the study of brain-damaged patients. However, cognitive neuropsychology is not a kind of neuropsychology. The former rests on the idea that a fruitful analysis of behaviour can be done *without* modelling the process in terms of brain mechanisms (Andrewes, 2001). Both disciplines differ from cognitive neuroscience in that the latter relies primarily on the study of normal subjects.
II. Anatomical and functional modularity

As discussed in the previous section, much of cognitive science is committed to the modular approach to the study of cognition. As we have seen, the core of this approach consists of an assumption concerning the basic structure of both the brain and the mind, what we have called the *modularity assumption*. It will now be useful to distinguish between two versions of this assumption. The first of these is the *functional* modularity assumption, which is the idea that the architecture of human cognition (or some part thereof) consists of a configuration of cognitive modules, where ‘module’ refers to a domain specific and functionally independent cognitive mechanism (Section I). What this means, in fact, is that human cognition can be decomposed into a number of functionally independent processes, and that each of these processes operates over a restricted domain of cognitive information. Moreover, since these processes are brain processes, to hypothesize that the capacity to do *A* and *B* depends on two distinct cognitive modules—one responsible for the capacity to do *A* and the other responsible for the capacity to do *B*—is to hypothesize that the brain processes cognitive information related to *A* *separately* from the way it processes cognitive information related to *B*.

It should be noted, however, that the nature of the processes composing two such cognitive modules may or may not be qualitatively different. What makes the *A* module/process distinct from the *B* module/process is their *functional independence*—i.e. the fact that one can be damaged, *in part or in totality*,...
without the other necessarily being damaged, and vice versa. So for instance, a pocket calculator could have four different division modules, one for dividing numbers smaller than or equal to 99 by numbers smaller than or equal to 99, a second one for dividing numbers smaller than or equal to 99 by numbers greater than 99, a third one for dividing numbers greater than 99 by numbers greater than 99, and a fourth one for dividing numbers greater than 99 by numbers smaller than or equal to 99. In such a calculator, these four capacities could all depend on (four versions of) the same algorithm. Yet, random damage to one of these modules would not necessarily result in the malfunctioning of the other modules.

The second version of the modularity assumption, the *anatomical* modularity assumption, is the idea that the cognitive modules which compose cognition (or at least most of them) each reside in a bounded and relatively small region of the brain. Understood in this way, the anatomical modularity assumption is the functional modularity assumption *plus* a claim about the implementation of functionally modular processes in the brain.

Notice that the level of generality with which both of these assumptions are stated makes them compatible with a wide range of possibilities as to what the specific nature of human cognitive architecture actually is. These modularity assumptions do not, for instance, tell us how many modules there are, nor do they tell us what kinds of cognitive functions are modular and what kinds aren’t. Moreover, these two assumptions do *not* constitute the only ways in which cognition could be modular, let alone the only ways in which cognition could be
organized. What makes these assumptions appear plausible to many cognitive scientists is the very existence of such disciplines as neuropsychology and cognitive neuropsychology. In the case of the functional modularity assumption, the presence of highly selective cognitive impairments like prosopagnosia (severe inability to recognize familiar faces with normal non-face object recognition) and various linguistic processing deficits (e.g. Broca’s aphasia) suggest the functional independence of at least some cognitive processes. Moreover, the fact that these highly selective impairments often arise in patients showing fairly localized brain lesions seems to provide considerable evidence for the anatomical modularity assumption. Some (e.g. Coltheart, 2001) further argue that the existence of neuropsychological research suggests not only that each of these assumptions seems justified, but also that both of them need to be true in order for that kind of research to succeed. Here, the reasoning is simply that if human cognition was characterized by overall functional modularity without overall anatomical modularity, we would expect almost any brain damage to result in multiple cognitive deficits, a situation that would render neuropsychology almost impossible to practice. Whether or not cognitive scientists are justified in making the two modularity assumptions is the subject of ongoing debate.

With these two assumptions in mind, cognitive psychologists often postulate the existence of cognitive modules on the basis of specific behavioral patterns. Let us call this general inferential strategy the functional modularity inference.

One of the most common behavioral patterns used in this sort of theorizing is a double dissociation of cognitive tasks. A cognitive task $A$ is said to be (singly)
dissociated from cognitive task $B$ when at least some individuals are observed who show a significant deficit with respect to $A$ in the absence of a corresponding deficit in $B$. $A$ and $B$ are said to *doubly* dissociated when, in addition, we observe individuals in whom $B$ is significantly impaired without a corresponding deficit with regard to $A$. Cognitive psychologists generally hold that dissociations are signs of separateness. If $A$ is observed to fail when $B$ does not, then one may infer that $A$ involves some process $M$ that $B$ does not involve, or at least that there is some process that $A$ more significantly draws on than $B$ does. When this process $M$ is obstructed, it is argued, $A$ fails and $B$ does not. In the case of a double dissociation, the inference is stronger, namely that $A$ and $B$ each involves (or significantly draws on) some process that the other does not. (See Shallice 1988 for a detailed discussion of this methodology). Where the process in question is a cognitive process, then, dissociations may provide evidence of the existence of cognitive modules.

Note that such behavioral patterns do not necessarily need to be observed in cases of brain injury. Behavioral dissociations can also be developmental. For example, it has been found that infants are able to differentiate large set-sizes on the basis of gross ratio, when that ratio is large enough, but not when it is small (Xu, Spelke, & Goddard, 2005). For example, they are able to differentiate between 8-membered sets and 16-membered sets, or between 16-membered sets and 32-membered sets, but not between 8-membered sets and 12-membered sets or 16-membered sets and 24-membered sets. Adults show analogous limitations on their ability to discriminate set-size for large sets.
However, adults are able to differentiate small sets even when the ratios are small—they can differentiate between 3-membered sets and 4-membered sets, for instance. By contrast, babies' failure to discriminate low-ratio sets extends to small-numbered sets as well. This observed dissociation between the ability to differentiate small sets in babies and adults is purely developmental.

Neuropsychologists and cognitive neuroscientists, for their part, often postulate the existence of anatomical modules (i.e. cognitive modules that are localized in bounded and relatively small brain regions) on the basis of the specific behavioral effects of brain lesions, as well as on the basis of various brain imaging and electrical recording data. Let us call this second general inferential strategy the anatomical modularity inference. In the next section, we will look at detailed examples of the use of both kinds of modularity inference.

Despite their overall appeal, the legitimacy of both the anatomical and functional modularity inferences has been strongly challenged for the past twenty years or so. In the case of the functional modularity inference, various authors have convincingly argued, on both theoretical and empirical grounds, that the existence of a double dissociation between subjects' performances on two different cognitive tasks does not necessarily constitute strong evidence for the existence of separate cognitive functions and/or modules (Dunn & Kirsner, 2003; Juola, & Plunkett, 2000; Plaut, 1995; Shallice, 1988; Uttal, 2001; Van Orden, Pennington, & Stone, 2001). For instance, double dissociations can be found in non-modular systems, i.e. systems that do not have causally or functionally independent components (Juola & Plunkett, 2000; Plaut, 1995; Shallice, 1988).
Similar arguments have also been proposed against the legitimacy of the anatomical modularity inference (Poeppel & Hickok, 2004; Uttal, 2001; Van Orden & Paap, 1997).

Such arguments, however, have had little or no impact on the ground. Indeed, the functional modularity inference continues to be one of the most common approaches among certain groups of cognitive psychologists for inquiring about the structure of cognition. Similarly, the recent cognitive neuroscience literature abounds more than ever with cases involving the use of the anatomical modularity inference. One possible reason for this is that the critics of the modular approach rarely suggest workable modifications to the current framework. In this respect, my dissertation can be viewed as an attempt to correct this situation.

In the next and final section of this chapter, I look at how the modular approach actually works in practice by considering the recent history of the face-recognition module hypothesis. In Chapter 2 and 3, following the introduction of an important, yet rarely made, distinction between two functional concepts—the distinction between working and use—I refocus the attention on new aspects of the main limitations of the anatomical and functional modularity inferences (Chapter 2). I then argue that these two kinds of inference can, and do, operate in three distinct modes in contemporary cognitive science, and that seeing this is essential to understanding both the power and the limitations of these methodological tools (Chapter 3). In this first part of the dissertation, I thus assume, along with modular theorists, that cognition is both functionally and
anatomically modular, and then proceed to identify and clarify some of the
important challenges to the modular approach.

In Part II (Chapters 4, 5, and 6), my goal is to develop a general approach to
the study of cognition, one that is in many ways an answer to several of the
challenges faced by the modular approach. What makes the approach appealing,
apart from being a genuine alternative to the modular approach, is that it favors
the coordination of research efforts within particular disciplines and across
disciplinary boundaries through the integration of theories, methods, and
assumptions. In Chapter 4, I introduce the approach, which I call the working
zone approach, and then show how it works on the ground using a recent
example from the cognitive neuroscience of working memory. In Chapter 5, I
highlight three distinct ways in which it brings integration in cognitive
neuroscience. The working zone approach 1- favors the integration of models
and theories within and across disciplines, 2- facilitates the use of functional
neuroimaging data to help advance cognitive theory, and 3- better accounts for
the continuity between animal and human cognition. Finally, In Chapter 6, I make
precise the working zone approach’s evolutionary commitments and argue that
these are better suited for the overall interdisciplinary project of cognitive
neuroscience than the adaptationist program that typically characterizes the
modular approach.
III. Case study 1: the face-recognition module hypothesis

Cognitive scientists make use of a wide variety of methods and techniques in imputing functions to structures. But amid this methodological diversity, it is interesting to note that all of them rely on what we might call *task analysis*, i.e. the decomposition of a cognitive task into a complex of more specific ones. In this respect, they all share a common methodological problem, namely, how to infer the functions of particular structures—or how to infer the organization of various cognitive functions—based on data derived from an analysis of people’s performance on particular cognitive tasks. This general methodological problem is at the center of the current debate over the use of both the anatomical and functional modularity inferences.

The capacity to recognize faces is one such general cognitive task which has been extensively studied by cognitive scientists. Moreover, several theorists have, over the past several decades, hypothesized the existence of a face-recognition module in the human visual system within the context of the modular approach. Thus, a good way to illustrate and come to grips with the way in which this general methodological problem manifests itself in the use of both modularity inferences is to consider the recent history of the face recognition research.

In its recent form, the face-recognition module hypothesis says that the cognitive systems responsible for face recognition are domain-specific (functional modularity) and that they lie in some specific and relatively small regions of the brain (anatomical modularity).
Several lines of evidence support the *functional* modularity of face recognition. In his seminal paper, Bodamer (1947), reflecting on a rare syndrome in which patients report being unable, on the basis of facial cues alone, to recognize highly familiar persons, labelled this deficit *prosopagnosia* to indicate that a face-specific process, distinct from the one for recognizing other objects, had been disrupted (Yin, 1970). Prosopagnosia is now defined in a more neutral way with respect to the exact nature of the underlying deficit. It is defined as a specific form of visual agnosia in which there is a deficit of facial recognition. This more neutral definition of the behavioral deficit is meant to capture the fact that a severe inability to recognize familiar faces almost always co-occurs with other recognition deficits, such as various forms of object agnosia (deficits in non-face object recognition) and other failures of visual recognition. Nevertheless, there are a few cases in the neuropsychological literature of patients who appear to have a recognition deficit specific for faces, sometimes referred to as *pure* prosopagnosia (see e.g. Rossion et al., 2003). Moreover, there are also reports of patients with impaired visual object recognition who retain face recognition (Rumiati & Humphreys, 1997). The presence of such a double dissociation between the capacity to recognise faces and the capacity to recognise objects is often believed to reflect the existence of (at least) two separate cognitive modules: a face-recognition module, and a visual-object-recognition module (Coltheart, 1999, p.119).

Evidence for the existence of a face-recognition module has also been produced within the context of the *anatomical* modularity inference. Both positron
emission tomography (PET) (Sergent & Signoret, 1992), and functional magnetic resonance (fMRI) (Kanwisher, McDermott, & Chun, 1997; McCarthy, Puce, Gore, & Allison, 1997) studies point to a relatively small region of the fusiform gyrus that responds more to pictures of faces than of other objects (even scrambled faces) in the normal human brain. On the basis of these findings, this region, which is now referred to as the fusiform face area (FFA), has been defined by some as a module specialized for face perception (Kanwisher et al., 1997).

The belief in the existence of a face-recognition module has had a significant impact on the development of theoretical models within various branches of cognitive science. A good example of this is Bruce & Young’s (1986) information-processing model of familiar face recognition. In this model, a lesion to the alleged face-recognition module, or to its connection with other modules such as the ‘person identity nodes’, is proposed as one of the main explanations of prosopagnosia. Similar models have also been proposed to account for Capgras Syndrome, a relatively rare disorder in which the patient believes that a close relative (most frequently a spouse) has been replaced by a nearly identical double, usually believed to be impostor (Breen, Caine, & Coltheart, 2000; Ellis & Lewis, 2001; Ellis & Young, 1990).

Despite its strong initial and current popularity among cognitive scientists, the idea that the existence of a face-specific system is what most plausibly accounts for both the characteristic breakdown of face recognition in prosopagnosia, and the particular patterns of brain activation under face recognition tasks, has never enjoyed universal support. Early criticisms pointed to the fact that in most cases
of prosopagnosia, recognition of familiar buildings, flowers, types of cars, or animal species may all be affected to varying degrees. This observation led to the alternative hypothesis that what accounts for prosopagnosia is an inability to recognize the individual members of categories that contain several items of similar appearance (Young, 1996).

Recent evidence in support of this ‘within-category’ (or ‘expertise’) hypothesis comes from an interesting case study by Dixon et al. (1998). The study involved the testing of one of their prosopagnosic patients (ELM) for his capacity to pair faces to names under two different conditions. In one of the tasks, he was asked to pair names to categorically related faces that shared multiple visual features and that were semantically close (e.g., female ice skaters, politicians, actors). In another task, ELM was asked to pair the faces taken from the same general pool, but this time with categorically unrelated people (e.g., singer Celine Dion, actress Betty Grable, and former First Lady Hillary Clinton). What was found is that ELM made tenfold fewer errors when faces were associated with semantically unrelated people. Moreover, the same pattern was also observed in the case of category-specific recognition problems involving viewing samples composed of nonface objects (fruits and vegetables, animals and birds, cars and stringed musical instruments). These results suggest that ‘prosopagnosia and co-occurring category-specific recognition problems both stem from difficulties disambiguating the stored representations of objects that share multiple visual features and refer to semantically close identities or concepts’ (Dixon et al., 1998, p.362).
This hypothesis is particularly interesting because it suggests that the behavioral differences between patients such as ELM and patients who seem to have pure prosopagnosia could reflect a variable degree of impairment to a unique cognitive process $P$. That faces can sometimes be the only category of objects to be affected could indeed reflect the fact that faces pose an even more difficult recognition problem than most other categories of objects. For one thing, faces must be identified at the level of the individual, whereas animals or fruits and vegetables are typically identified using basic-level categories (e.g. tiger or apple). In addition, whereas the members of most categories of objects which typically pose problems for prosopagnosics (e.g. categories of biological objects) can be distinguished from related members using the presence or absence of unique features (e.g. colors in the case of fruits and vegetables, stripes in the case of animals), most faces must be distinguished using relatively small differences in features that all members of the category possess (Dixon et al., 1998). Thus, minor impairment to $P$ could result in behavioral effects characteristic of pure prosopagnosia, while more substantial degrees of impairment could result in behavioral effects characteristic of ELM.

Additional evidence in support of this 'one process' hypothesis comes from a recent study by Gauthier et al. (2003) which directly addresses the issue of the functional independence of face and object processing. Using behavioral measures of interference, what was found is that holistic perception (obligatory processing of all parts) for cars interferes with concurrent holistic perception of faces by car experts, but not by car novices. Moreover, electrophysiological
measures of this interference indicated that face and car holistic processing reflect the activity of the same neural network. This finding is especially important since it provides *direct* evidence against the hypothesis of functional independence between the mechanisms responsible for face and object holistic perception.

Furthermore, several neuroimaging studies have seriously challenged the FFA (i.e. anatomical) version of the face-recognition module hypothesis. First, an expertise effect (increased activation) in the FFA was reported with unfamiliar nonface objects (Gauthier, Tarr, Anderson, Skudlarski, & Gore, 1999). Second, a subsequent study extended these results to other homogeneous categories of objects by showing similar expertise effects for the FFA in car and bird experts (Gauthier, Skudlarski, Gore, & Anderson, 2000). Together, these results suggest that ‘the level of categorization and expertise, rather than superficial properties of objects, determine the specialization of the FFA’ (Gauthier et al., 2000, p. 191. See also Palmeri & Gauthier, 2004 for a review of this literature).
The origin and recent development of the face-recognition module hypothesis brings into full light the general methodological problem, mentioned earlier, of inferring the function of a particular cognitive structure—or inferring the architecture of a particular cognitive function—on the basis of an analysis of subjects’ performance on various cognitive tasks. This problem arises quite sharply in disciplines such as cognitive (neuro)psychology, neuropsychology, cognitive neuroscience and evolutionary psychology, where the modularity approach enjoys a fair amount of popularity.

A primary concern, in this context, is that the practice of inferring cognitive modules via the anatomical and functional modularity inferences may often be too detached from implementation concerns for it to fulfill its goal of identifying the basic components of cognitive systems. The concern here is that when considered on their own, behavioral data, such as a double dissociation between object and face recognition, or functional neuroimaging data, such as the selective activation of the FFA for faces, only provide weak support for the existence of distinct cognitive modules. Generally stated, the main reason for this lies in the observation, noted in section II of the previous chapter, that very different kinds of cognitive architectures (e.g. modular and non-modular architectures) may give rise to similar behavioral and brain activation patterns.
(e.g. double dissociations and the selective activation of brain regions for particular classes of stimuli).

In this chapter and the next, my goal is to further clarify this concern by offering a new characterization of the anatomical and functional modularity inferences. To do this, I introduce an important, yet rarely made distinction, between two functional concepts—the distinction between working and use (section I). I then proceed with a new characterization of the modularity inferences by first retracing their roots in the development of modern neuropsychology in the second half of the nineteenth century (section II and III). In Chapter 3, I continue my analysis by arguing that the modularity inferences can, and do, operate in three distinct modes in contemporary cognitive science, and that seeing this is essential to understanding the power and limitations of these methodological tools.

I. Working vs. use: two distinct modes of functional specification

The functional specification of cognitive structures can operate in two importantly different modes, which correspond to two different senses of the term ‘function’. In the first mode, the ‘function’ of a cognitive structure is specified in terms of a cognitive operation, or set of such operations, that the structure is thought to perform (i.e. carry out internally). In this case, what is specified is the structure’s cognitive working. In the second mode, the ‘function’ of a cognitive structure is
specified in terms of a cognitive operation, or set of such operations, in which the structure is thought to take part, or participate. In this case, what is specified is the structure’s cognitive use.

A simple example from physiology will serve to illustrate these two senses of the term ‘function’, as well as the two corresponding modes of functional specification. The circulation of blood and the performance of rhythmic contractions are two physiological ‘functions’ of the heart. On the one hand, the circulation of blood is an activity in which the heart participates, but it is not an activity that it performs (it is the cardiovascular system that performs this activity). The circulation of blood is thus specified as a physiological use of the heart. On the other hand, the heart performs rhythmic contractions, and this activity is thus specified as a physiological working of this organ.

The same distinction now applies in the case of cognitive ‘functions’. For example, we know that several areas of the brain contribute to face recognition (Haxby, Hoffman, & Gobbini, 2000; Rossion et al., 2003). In this respect, face recognition is a cognitive activity in which the fusiform face area (FFA) participates, not a cognitive activity that it performs. Face recognition is thus specified as a cognitive use, not as a cognitive working, of the FFA.

As for an example of cognitive working, consider the computational model of visual word recognition and reading aloud proposed by Coltheart et al. (2001). The model, which the authors refer to as the dual route cascaded (DRC) model, consists of three cognitive routes via which the normal capacity to read aloud is made possible. Fig. 1 below depicts the two lexical routes and the non-lexical
route proposed by the model. Take for instance the grapheme-phoneme rule system box, which is a part of the non-lexical route. This box represents a cognitive subsystem which implements a set of rules for converting graphemic representations (held in the letter units system) into phonemic representations (held in the phoneme system). Notice that the functional specification of this cognitive component is done relative to the cognitive operation that it is thought to perform, namely, the conversion of graphemic representations into phonemic representations. This cognitive operation is thus specified here as a cognitive working of this cognitive component.

Fig. 1. Coltheart et al. (2001) dual route cascaded model of visual word recognition and reading aloud. See text for details.
Note the following four points. Firstly, the distinction between working and use is an ontological distinction, and not merely a distinction of characterization. For example, in the case of the heart, the distinction between rhythmic contraction and circulation of blood is an ontological working-use distinction, because the circulation of blood is something that the heart does in conjunction with other things. However, the distinction between pumping blood and rhythmic contraction is a characterization-distinction, because these are two ways of characterizing the heart’s own activity.

Secondly, the distinction between working and use is not a distinction between levels of description. Depending on which level of description is most suitable in any given context, both the working(s) and the use(s) of a cognitive structure can be described at various levels of analysis ranging from the neurobiological to the functional level. For example, Loffler et al. (2005) hypothesize that individual human faces are represented in the FFA ‘by their direction (facial identity) and distance (distinctiveness) from a prototypical (mean) face’ (p. 1386)—a purely cognitive description of the FFA’s working—while Jiang et al. (2006) hypothesize that individual human faces are represented ‘by neurons tuned to face shapes that are located in the FFA’ (p. 159)—a neurocognitive description of the FFA’s working. These hypotheses concern the working (as opposed to the use) of the FFA because they are meant to capture cognitive activity (representation) that occurs within this structure.

Thirdly, what makes a particular functional label refer to a cognitive working as opposed to referring to a cognitive use, or vice versa, depends on the particular
brain region or cognitive component it is attributed to. In other words, functional claims such as ‘X is a cognitive working’, or ‘X is a cognitive use’, where X is a functional label, are incomplete. Such claims must always take the form ‘X is a cognitive working of p’, or ‘X is a cognitive use of p’, where p is a brain region or cognitive component. A consequence of this is that a particular functional label can refer to a cognitive use in one context, while referring to a cognitive working in a different context. For example, while the functional label ‘face recognition’ refers to a cognitive use of the FFA, the same label may refer to a cognitive working of some network of brain regions N (which will include the FFA).

Fourthly, the word ‘use’ is not intended in a normative sense. An activity that has no evolutionary value to the organism, or which is even detrimental to it, might still count as a use in my sense. For instance, it might be that FFA is used to recognize individual cars. This is a use even though FFA was not selected for this. Further, it might be that this use of FFA is actually detrimental to the human user—maybe it interferes with face-recognition in certain contexts. It is still a use in my terminology. What uses a cognitive structure has must therefore be determined completely independently of their particular value to the organism. True, we may expect that the presence of at least some cognitive structures will be explained by the historical value of their uses to organism which possesses them—e.g. cognitive structures with historically beneficial uses would have been the object of natural selection—but the point is that an analysis of such historical (or current) value is not necessary to determine what particular uses a cognitive structure has at any given point in time (more about this in Chapter 6).
II. A new perspective on the modularity inferences

Having distinguished between cognitive workings and cognitive uses, we are now in a position to say something a little more precise about the nature of the anatomical and functional modularity inferences.

Starting with the functional modularity inference, we can first point out that this inference typically rests on a set of data composed mainly, if not exclusively, of dissociation data of various kinds (e.g. dissociations observed in developmental, comparative, or brain lesion studies). These data, as we have seen, derive from an analysis of the performances of normal or brain damaged subjects on various cognitive tasks. As such, they consist of the description and specification of the behavioral consequences of the (mal)workings of a cognitive component (or of several different cognitive components). More specifically, dissociation data consist of a description and specification of the cognitive effects that the (mal)workings of a cognitive component has on the cognitive operations in which that component participates. Consequently, these essentially behavioral data will most likely be expressed in ways that capture one or more of the cognitive uses of a component (or group of such components) within the larger system. In other words, relative to the component of interest (e.g. a damaged structure), dissociation data will most often consist of the description and specification of (some of) that component’s cognitive use(s), and will only very rarely consist of an adequate description and specification of its cognitive working(s). This, in turn, means that a set of dissociation data is unlikely to tell us
much about the structure of the cognitive operation under investigation, or about the internal organization of the cognitive system(s) on which it runs. (For example, a double dissociation between face- and object-recognition does not tell us whether faces and objects are processed by separate systems, or whether face- and object-recognition share most, or only a few, of their components.)

The same reasoning can be applied in the case of the anatomical modularity inference, which in fact consists of the functional modularity inference plus a claim about where in the brain these modular processes occur. In this case, the dissociation data are not purely behavioral. As in the case of the functional modularity inference, the dissociation data are derived from an analysis of the performances of subjects on various cognitive tasks, but these behavioral patterns are then linked to various anatomical data. In the case of neuropsychology, the behavioral dissociations are linked to specific brain lesions, and in the case of cognitive neuroscience, the behavioral patterns are linked to various kinds of recordings of brain activity.

The foregoing discussion suggests that in order for the anatomical and functional modularity inferences to be justified, one must secure the transition from the specification of cognitive uses—derived from behavioral data—to the specification of cognitive workings—required in order adequately to understand the internal organization of cognitive systems. It is important to recognize hypotheses that posit cognitive uses and those involving cognitive workings, and to be aware of the different kinds of evidence that are able to give these appropriate support. Let us therefore specify the goal of the modularity
inferences as an attempt to secure the transition from the specification of cognitive uses to the specification of cognitive workings.

In the light of this particular characterization of the modularity inferences, I would now like to make more precise what I think is one of the most serious problems with how these methodological tools are currently being used. In a nutshell, this problem is that the modularity inferences often lose much of their force (or are greatly weakened) because modules are specified at too high a level. In fact, this problem arises mainly from a failure to adequately secure the transition from the specification of cognitive uses to the specification of cognitive workings. In support of these claims, I consider next two cases of the application of the anatomical and functional modularity inferences, and show that my analysis can shed some light in each of these cases on how best to evaluate the strength of the inferences. The first case concerns the development of neurolinguistics in the second half of the nineteenth century, and the second one is taken from the field of visual object recognition.

III. Case study 2: Wernicke’s insight

The French physician Jean-Baptiste Bouillaud was apparently the first to apply the anatomical modularity inference to the localization of higher cognitive functions. His 1825 paper on the localization of the speech articulation organ in the frontal lobes contains a clear explanation of the logic of the strict clinical-
pathological correlation method (Bouillaud, 1825). The latter, which included a version of what later was to be termed ‘double dissociation of function’ (Teuber, 1955), consisted in the localization of specific psychological faculties by observation of specifically impaired capacities following focal brain damage (Graves, 1997). Bouillaud’s contribution to the methodology of discovering specialized functional areas of the human brain had its first significant impact on the development of neuropsychology when Paul Broca (1861; 1865) made use of the same paradigm to infer that the function of speech articulation was located in the posterior third frontal gyrus of the left hemisphere (Graves, 1997).

Characteristic of Bouillaud’s method, as it was applied by Broca, was the postulation of specific, fairly isolated, and self-sufficient centers for each function that could be discretely impaired by a focal lesion. This, however, did not provide an account of how higher psychological functions emerge from more simple (or basic) ones. Moreover, the domain specificity of these functional centers (modules) was simply determined on the basis of the specificity of the observed behavioral deficits.

The current modular approach to the study of cognition has retained at least two important aspects of Bouillaud’s method as it was applied in the early days of neurolinguistics. The first of these aspects is a strong commitment to some version of both the anatomical and functional modularity assumptions. In the nineteenth century, the widespread adoption of these assumptions reflected primarily the strong influence of Franz Joseph Gall’s localizationist views.5

5 Bouillaud in fact accepted Gall’s principle of radical localization of higher faculties, but he replaced Gall’s phrenological methodology with the clinical-pathological correlation method.
Today, as we have seen, such a commitment finds its justification in the very existence of neuropsychological research.

The second aspect of Bouillaud’s method that is preserved in the current framework is that in most cases, the domain specificity of the hypothesized centers (modules), i.e. their functional specification, is often derived exclusively from the available pool of specific cognitive deficits. In this way, the level at which the patients’ performances on various cognitive tasks are specified is often what determines the level at which the specification of these centers will occur.

In 1874, with the publication of Wernicke’s original monograph, the study of aphasia entered a new era. Wernicke’s new approach to localization stood in sharp contrast to the schools that were dominant at the time. This fact is often obscured in recent discussions, due to the fact that one of Wernicke’s most memorable contributions to the field was the localization of a sensory speech area, which bears his name, and which is sometimes (mistakenly) referred to as the language comprehension area. Wernicke was in fact very resistant to localizing any cerebral centers beyond what he referred to as the ‘primary’ (motor and perceptual) ‘psychic functions’. His basic model of the aphasia symptom complex was limited to the identification of only two centers, the ‘acoustic images’ centre, and the ‘motor speech images’ center. But the most striking feature of Wernicke’s model is not that it contains a very small number of centers. What makes his approach radically different from that of his predecessors is that it involves the elaboration of a new structure/function paradigm. While the application of Bouillaud’s method by Broca and his followers
consisted in an attempt to link higher functions (e.g. speech articulation, reading) to specific anatomical centers in the brain, Wernicke’s strategy focused instead on relating these higher capacities to underlying sensory and motor projection systems. The following quote from Wernicke’s second monograph nicely illustrates this different focus.

In my first work on aphasia I took pains to show that in such an interpretation of the speech process as has been reviewed above, we had probably found the scheme of cortical function as a whole, that memory images were the psychic elements populating the cortex in a mosaic-like arrangement as a functional development, which may very well be localized according to the regions of the nerve-endings, so that the acoustic images find their abode within the cortical terminals of the acoustic nerves; the visual images, within the cortical endings of the optic nerve; and the olfactory images in that of the olfactory nerve and so on. Likewise, the motor memory images or movement-representation could be located in the cortical sites of the motor nerve origins. For example, the images of speech movements would then be found in the Broca gyrus and those of writing in the cortical area serving arm movements, etc… *Any higher psychic process, exceeding these mere primary assumptions, could not I reasoned, be localized, but rested on the mutual interaction of these fundamental psychic*
elements mediated by means of their manifold connections via the association fibers (Wernicke, 1885) (Wernicke, 1885-1886/1977, pp.177-8, emphasis added).

In Wernicke’s model, only elementary sensory and motor functions are held to be both functionally distinct and discretely localized. All other (higher) psychological functions (e.g. speech production and language comprehension) are composed of the interactions between these basic functions, and thus all higher functions are held to be both functionally and anatomically interrelated. This decomposition of higher functions into more basic principles and functions made his model theoretically explanatory, and it also allowed for the successful prediction of new syndromes.

The development of Wernicke’s model is regarded by some as the most productive and significant period in the history of aphasia. With Lichtheim’s subsequent systematization (Lichtheim, 1885), the Wernicke-Lichtheim model quickly became the standard neuropsychological theory. Thanks to the rise of behaviorism, the first half of the 20th century saw a sharp decrease in its popularity until Norman Geschwind revived it in the 60s (Geschwind, 1965). Today, many of the core components and concepts originally proposed by Wernicke are finding their way into the emergent conceptual and empirical trends that form the new functional anatomy of language (Poeppel & Hickok, 2004), and his 1874 model is still taught across the various language sciences. Such congruence with much of contemporary thought on brain function and how it
relates to the complex symptomatology of the aphasias points to the value, for contemporary theorists, of engaging in a re-examination of Wernicke’s original works. But more importantly for us here, it brings to light the fact that the development of Wernicke’s model was one of the first (and perhaps the most) successful application of what we have called the modular approach to the study of cognition.\textsuperscript{6} An appreciation of this fact requires that we examine the development of the model in more details.

A reading of the 1874 monograph reveals the complex set of factors which led to the postulation of Wernicke’s original model.\textsuperscript{7} The latter was founded on the application of Theodor von Meynert’s general theory of cortical function. The central tenets of this theory were that cortex that lies anterior to the Rolandic sulcus serves motor functions, whereas cortex that lies posterior to this sulcus serves sensory functions, and that the two areas are linked by the presence of a fiber tracts system, the association tracts. The significance of the adoption of this general scheme can be appreciated by considering the fact that it was the introduction of such a sensory-motor dichotomy in the organization of the brain that appears to have been at the origin of the concept of receptive (or sensory) aphasia. In fact, as Whitaker & Etlinger (1993) recently pointed out, one of Wernicke’s most significant achievements in the 1874 monograph was ‘the

\textsuperscript{6} As can be seen from the quote above taken from his second monograph, Wernicke’s approach – or, more precisely, his particular take on the anatomical and functional modularity assumptions – differs greatly from the approach that was introduced at the outset as the ‘modular approach to the study of cognition’. The latter, which is often referred to as the \textit{massive modularity thesis}, is thought to be applicable to all (or at least most) domains of cognition. In this context, Wernicke’s approach appears to deserve the name of \textit{partial modularity thesis}.

\textsuperscript{7} The following brief review of the main factors at work in the development of Wernicke’s model is largely based on G. H. Eggert’s English translation of Wernicke’s first monograph (Wernicke 1874/1977, pp.91-145).
reconstruction of the idea of language in such a way that comprehension or reception was now an individual component of language’ (p.562). Prior to such conceptualization, language was conceived as an essentially productive or expressive power. A receptive capacity such as language comprehension (or understanding) was simply not part of what Wernicke’s predecessors, including Gall and Broca, called language. Whitaker & Etlinger (1993) then conclude that language comprehension had thus become ‘a candidate for cerebral localization’ (p.562). This is not quite correct. Language comprehension is a complex faculty that Wernicke did not think could be localized, as mentioned above.

But in order to support his theoretical model, Wernicke had to demonstrate the existence of this new language component through the identification of the associated syndrome, namely, receptive aphasia. Moreover, the anatomical and functional independence which is characteristic of the acoustic and motor speech images centers demanded that the two corresponding forms of aphasia be shown to have the same degree of independence, both functionally as well as anatomically. Evidence for the functional independence of these centers ultimately came from a few of Wernicke’s own cases, and it relied on the same dissociation logic that Bouillaud had described many decades earlier. Thus:

The great variability of the clinical picture of aphasia moves between the two extremes of pure motor aphasia and the pure sensory form. The demonstration of these two types must be regarded as conclusive proof of the existence of two anatomically

This passage clearly indicates Wernicke’s reliance on what we have called the functional modularity inference, for which the behavioral data consist, in this case, of a double dissociation between pure motor aphasia and pure sensory aphasia. In that respect, there is nothing new to Wernicke’s methodology. As we mentioned earlier, the same kind of reasoning had been applied for quite some time by the proponents of the strict clinical-pathological correlation method, of which Broca was probably the most successful. But what is radically different in Wernicke’s approach is that, unlike his 19th century predecessors, his functional specification of both language centers did not derive directly from his specification of the observed syndromes. The latter, which included symptoms such as confused but fluent speech pattern, loss of language comprehension, and loss of speech production, were all thought by Wernicke to arise from an abnormality in the functioning of one or more of the basic psychic (perceptual and motor) elements. It is thus at this basic functional level that he thought the language centers had to be specified. And since Wernicke’s knowledge of the functional and structural organization of these basic psychic elements had come mostly from his understanding of brain anatomy and physiology, it was this particular knowledge that formed the basis of his functional specification of the acoustic and motor speech images centers. The discovery of a double dissociation between speech production and language comprehension did have
an important role to play in the development and justification of the model, but as the last quote from Wernicke indicates, that role consisted in providing evidence for the existence of (at least) two separate centers, not for the existence of any particular center, let alone for the existence of a speech production and/or a language comprehension center.

The 1874 work also included Wernicke’s hypothesis about the localization of the acoustic images center on the basis of pathological findings of the destruction of the left superior temporal area in certain cases of aphasia, including one of Wernicke’s own patients which he described as a pure sensory aphasic. There is no doubt here that Wernicke was relying on what we have called the anatomical modularity inference.

In the light of all this, I would like to make the following conjecture: Wernicke’s successful application of the modular approach to the study of language can be attributed largely to his concern with the specification of cognitive workings (as opposed to the specification of cognitive uses). Wernicke himself did not make explicit the distinction between the cognitive workings and the cognitive uses of the various components of his model. However, it is clear from his writings that he had a good grasp of the main idea behind such a distinction, and that his functional specification of the model’s components was intended to reflect their cognitive workings, not their cognitive uses. That this was the case is perhaps best shown by his careful use of the evidence he produced for the existence of a pure form of both sensory and motor aphasia. In the hands of Broca, the same evidence would have almost certainly resulted in a very different picture of the
functional anatomy of language, with the most likely scenario being the hypothesis of a language comprehension center to complement the already hypothesized speech articulation center.⁸ There are certainly many possible factors that could have contributed to this (hypothetical) situation, but there is one which appears to be especially significant. While both Wernicke and Broca subscribed to some version of what we have called the anatomical and functional modularity assumptions, Wernicke’s attitude toward these assumptions was more conservative, i.e. more cautious. Indeed, while Broca had shown a strong commitment to Gall’s principle of radical localization of higher faculties, Wernicke’s understanding of the big picture of the mind was more systemic and interactive, due to the influence that Meynert’s general theory of cortical function had on his theorizing.

In retrospect, Wernicke’s attitude toward the modularity assumptions might appear to have been a little too cautious. It is now uncontroversial that his two-center model founded on a mosaic of more or less homogeneous sets of sensory and motor images cannot account for our current understanding of aphasic symptomatology. For instance, the presence of agrammatism in the speech output of some aphasics, the existence of anomic aphasia, and the fact that Broca’s aphasics typically have mild sentence-level comprehension deficits, to name only these three, suggest a more complex architecture than the one suggested by Wernicke’s model (Poeppel & Hickok, 2004). In fact, Wernicke himself was perfectly aware that his model of language was underspecified.

⁸ Graves (1997) makes a similar claim, though on somewhat different grounds.
Toward the end of the 1874 monograph, he writes:

In dealing with this problem [whether the use of still inconclusive anatomic and physiological studies is justified in the construction of a new theory], I have never succumbed to the temptation to go beyond the basic elementary hypothesis which now can hardly be seriously challenged, that the central nerve endings are invested with the role of psychic elements. Thus many of the oddities in the province of aphasia must remain unsolved, such as the isolated loss of substantives or verbs, etc. (Wernicke, 1874/1977, p.143).

Wernicke’s strategy seems to have been to proceed from the hypothesis of centers with the most basic (i.e. general) workings and domains of operation towards the hypothesis of centers with progressively more specialized workings with more specific domains of operation. This stands in sharp contrast to much of the current application of both the anatomical and functional modularity inferences in various fields of cognitive science.

Consider, for example, the field of visual recognition. A review of the recent literature indicates that a dominant trend among cognitive psychologists, neuropsychologists, and cognitive neuroscientists seems to have been to proceed in the opposite direction. This is particularly true in the case of face recognition, as we have seen in section 1, but it is also true more generally, as we will now see, of the field of visual object recognition.
Several neuropsychological reports point to the existence of category-specific deficits in human visual recognition. In addition to the double dissociation between face and object recognition described earlier, the recognition of natural objects (e.g. living things such as animals and plants) and manufactured objects (e.g. non-living things such as various tools and utensils) also appear to be doubly dissociable (see Warrington & Shallice, 1984 for reports of patients with a deficit specific to natural object recognition, and ; Warrington & McCarthy, 1994 for the opposite deficit). The existence of these specific deficits, or agnosias, have provided the motivation for a number of recent imaging experiments that have reported discrete cortical regions in the ventral temporal cortex of humans that respond preferentially to category specific stimuli including faces (Kanwisher, McDermott, & Chun, 1997; McCarthy, Puce, Gore, & Allison, 1997), letters (Polk & Farah, 1998), animals and tools (Chao, Haxby, & Martin, 1999), and buildings (Ishai, Ungerleider, Martin, & Haxby, 2000; see Joseph, 2001 for a review of this literature). Together, these findings have been interpreted by some (e.g. Kanwisher, Downing, Epstein, & Kourtzi, 2001) as evidence for a category-specific and anatomically segregated modular organization of object recognition in this region of the brain.

An alternative possibility, however, is that the representation of objects in the ventral visual pathway is more widely distributed, and that its organization is feature- rather than object-based (Ishai, Ungerleider, Martin, Schouten, & Haxby, 1999). This alternative hypothesis is more in line with physiological results in monkeys (Tanaka, 1993; 2003). Moreover, this proposal is consistent with the
idea that in the course of mammalian evolution, the capacity to recognize new categories of objects could result from the differential recruitment and/or enlargement of (preexisting) brain regions with expertise in the detection of particular features, a suggestion which is in line with emerging theories of brain evolution (Allman, 1999; Finlay, Darlington, & Nicastro, 2001).

But again, what is perhaps the most significant difference between these two proposals is that they involve quite different handlings of both modularity inferences. In the case of the object-based hypothesis of object recognition, as in Bouillaud and Broca’s application of the strict clinical-pathological correlation method, the functional specification of the postulated modules derives directly (and entirely) from the specification of the observed dissociations. In the case of the feature-based hypothesis, as in Wernicke’s model, the existence of these dissociations is used instead to infer the existence of functionally independent cognitive structures. The functional specification of these cognitive structures is then made on the basis of additional extraneuropsychological or extraneuroimaging data and hypotheses, including results from (but not restricted to) human brain anatomy, comparative neuroanatomy, neurophysiology, developmental psychology and evolutionary biology (more about this point in Chapter 5). The reliance on such additional hypotheses, as well as a more cautious handling of the modularity inferences, are, I would like to suggest, what allowed Wernicke to operate the modularity inferences at the appropriate level – i.e. the level at which cognition appeared likely to be modular given the state of neuroscientific knowledge at that time – which is in turn what made for the
impressive success of his model.
In Chapter 2, I argue that there is a difference between modularity inferences involving cognitive uses and those involving cognitive workings. This helps us understand one of the main challenges to current uses of these methodological tools. In the present chapter, my goal is to push the analysis further by describing three distinct modes in which the modularity inferences can, and do in fact, operate in contemporary cognitive science. I also discuss an ambiguity in the notion of cognitive module that is brought to light by my analysis.

I. Three operating modes

The two case studies of Chapters 1 and 2 provide an account of the origin, early development, and current uses of the modularity inferences in cognitive science. As these cases indicate, the anatomical and functional modularity inferences are flawed when understood in a certain way. In a way, their logical structure is analogous to the fallacy of affirming the consequent. In this case, it is first observed (correctly) that specific lesions to modular architectures (M) give rise to double dissociations (DD) between cognitive tasks (or that modular architectures
give rise to the selective activation (SA) of brain regions for specific cognitive
tasks.) Then, when a particular double dissociation is observed between two or
more cognitive tasks (or when the selective activation of particular brain regions
is observed for specific cognitive tasks), one infers that the cognitive system of
interest must have a modular architecture. Formally we have:

\[ M \rightarrow (DD \lor SA) \]
\[ DD \lor SA \]
\[ \hline \]
\[ M \]

The problem with this inference is that modular architectures are not the only
kind of architecture that can give rise to double dissociations of cognitive tasks
(or to the selective activation of brain regions) (Juola, & Plunkett, 2000; Plaut,
1995; Shallice, 1988; Van Orden & Paap, 1997). This is represented by the
single arrow in the first line of the formal inference. If modular architectures were
the only kind of architecture that could give rise to double dissociations of
cognitive tasks, or to the selective activation of brain regions, then the first line of
the formal inference would have a double arrow, and the inference would
therefore be valid.

That the modularity inferences have this logical form does not mean, however,
that they are necessarily useless or scientifically illegitimate. In fact, the
modularity inferences can be described as *inferences to the best explanation*,
and just like other kinds of non-deductive scientific inferences (e.g. analogical and statistical inferences), this kind of inference does play an important role in science (see for instance the helpful discussion in Lipton, 2004). In this case, one would conclude that a cognitive system must have a modular architecture if this hypothesis best explained an observed double dissociation (or pattern of brain activation) in that system, given the fact that such architectures are known to produce such behavioral and brain activation patterns. In all such cases, the proposed explanation would be contrasted against other plausible explanations. In the neuropsychological case, we know that some lesions to non-modular architectures can also produce double dissociations (Shallice, 1988). And in the neuroimaging case, we know that the selective activation of a brain region for a particular cognitive task does not rule out the possibility that it be also activated by other yet untested cognitive tasks. The less plausible these alternative explanations appear to be in any given case, the stronger the modularity inference is.

Variant assessments of the plausibility of alternative explanations account for the very different ways in which theorists handle DD and SA data. It is what accounts for the fact that Wernicke’s application of the anatomical and functional modularity inferences might have been radically different from Broca’s application, and the fact that the current use of these tools in the field of visual recognition, and especially face recognition, varies so drastically among different theorists. We can in fact identify three distinct modes in which the modularity inferences can, and do, operate in contemporary cognitive science. Let us
introduce them in turn.

The first mode is well represented by Broca’s application of Bouillaud’s strict clinical-pathological correlation method (discussed in Chapter 2 section III) to infer that the faculty of speech articulation was located in the posterior third frontal gyrus of the left hemisphere. It is also represented by the use of a double dissociation between various object and face recognition tasks as the main piece of evidence in support of the claim that beyond early visual processing stages, faces are processed by a separate cognitive system dedicated to faces (De Haan, 2001). This mode is by far the most ambitious use one can make of the modularity inferences, but it is also where the inference is the weakest. In this case, the functional or anatomical modularity inference can be seen as a combination of two claims.

The first of these claims consists of specifying the working (as opposed to the use) of the hypothesized cognitive component. It consists of specifying precisely the cognitive operation(s) that the hypothesized component carries out internally (e.g. the measurement or identification of some visual feature). Let us call this the working specificity claim. The second of these claims consists of specifying the component’s informational domain (e.g. the above operations are used exclusively if face recognition). Let us call this the domain specificity claim. Thus, to infer that beyond early visual processing stages faces are processed by a separate cognitive component dedicated to faces is 1- to infer that face recognition is a working, and not simply a use, of the hypothesized component
(working specificity claim), and 2- to infer that the component’s working is restricted to faces (domain specificity claim). In other words, it is to infer that the human capacity to recognize faces is performed by the working of a domain-specific module dedicated to faces. Accordingly, adequate evidence for both of these claims must be provided if the use of this operating mode is to be justified.

The problem with this mode is that in the vast majority of cases, the bulk of the evidence supports only one of the two claims. In fact, it is the second claim—the specification of the component’s domain of operation—which is typically the focus of attention. For example, most of the evidence currently available for the existence of a face-recognition module (as defined above) supports the existence of a face-specific component (domain specificity claim), not that face recognition is a working (as opposed to a use) of that component (working specificity claim). This brings us to the second mode.

The second operating mode of the modularity inferences is much less ambitious than the first one. It is also best described in relation to the first mode. The latter as been described as a combination of two claims, one being the specification of the hypothesized component’s working, and the other being the specification of that component’s domain of operation. The second operating mode of the modularity inferences consists of specifying the hypothesized component’s domain of operation without specifying that component’s working (i.e. only specifying its use). In other words, it involves making the domain specificity claim without making the working specificity claim.

An example of the functional version of this mode is the use of dissociation
data between various object and face recognition tasks to infer the existence a face-specific cognitive component. Notice here that this claim says nothing about the specific nature of that component’s internal operation (i.e. its working specificity). It says that the capacity to recognize faces can be decomposed into a number of contributory capacities—i.e. that there is functional specialization in the visual system—and that at least one of these contributory capacities, whatever this capacity turns out to be, is accomplished by the working of a face-specific component.

An example of the anatomical version of this mode is the claim by Kanwisher et al. (1997) that the FFA constitutes a “module specialized for face perception” (see Chapter 1, section III). This claim can be broken down into the following three observations:

1. Neuroscience can be taken to have shown that a face-specific cognitive component is localized in a bounded and relatively small region of the brain, namely, the FFA. This does not mean, of course, that all data-processes relevant to face perception are confined to this region. Some may involve the interaction of many widely distributed components passing data back and forth.

2. The neuroimaging data do not show that there is no overlap between face and object perception. What they suggest is that there is an area of non-
overlap between the two processes.

3. The claim, therefore, should not be taken to imply that face perception is localized in the FFA. In fact, Kanwisher et al. (1997) emphasize a number of other areas that are sensitive to faces.

Being less ambitious than the first mode, this mode appears to be the most popular of the two among cognitive scientists. A good indication that this is indeed the case is, as mentioned earlier, that the bulk of the evidence provided in favor of the existence of cognitive modules typically supports the domain specificity claim, but not the working specificity claim. Accordingly, the typical criticism leveled against the second operating mode of the modularity inferences is to challenge the domain specificity claim. A striking example of this has been the extensive debate in the cognitive neuroscience literature regarding the status of the FFA as a face-recognition module (see Chapter 1, section III). This debate has been completely dominated by the question of whether the FFA is face-specific, and it is only recently that investigators have started to tackled the question of what specific operation(s) does the FFA actually carry out (e.g. Jiang et al., 2006).

The third operating mode of the modularity inferences is even less ambitious than the second one. In this case, neither the working specificity claim nor the domain specificity claim is made. What is inferred is the functional independence,
or some degree thereof, of two or more cognitive operations. This is done by inferring the existence of distinct cognitive components with restricted, yet unspecified, domains of operation. For example, after observing a double dissociation between cognitive tasks A and B, one is led to infer that there is at least one distinct cognitive component that is necessary for performing task A but not for performing task B and vice versa. Similarly, after observing the selective activation of a brain region R for cognitive task C, one is led to infer that region R is the site of a cognitive operation that is necessary for the performance of task C but not for other tested tasks.

This form of reasoning is well represented by Wernicke’s application of Bouillaud’s strict clinical-pathological correlation method to infer the existence of a new brain center involved in language comprehension (see Chapter 2, section III). In this case, as we have seen, Wernicke did not use his discovery of a double dissociation between speech production and language comprehension to infer the existence of a center specific to language comprehension (as in the second mode), let alone to localize the faculty of language comprehension in a specific region of the brain (as in the first mode). Instead, Wernicke made use of the dissociation data to infer that the new center played a crucial role in language comprehension—i.e. that one of the important uses of the new center was language comprehension—leaving it open as to whether it was specific to this capacity.

Wernicke actually called the new center the ‘acoustic images center’. This particular specification of the center’s working and domain of operation was
ultimately based on his knowledge of functional anatomy and the location of his patients’ brain lesions. The main lesions corresponding to the language comprehension deficit were located in the area adjacent to the auditory nerve ending at the junction of the temporal and parietal lobes. Wernicke was thus very careful not to specify the new center’s working based solely on his specification of the patients’ behavioral deficits, thereby avoiding the pitfalls associated with the first operating mode of the modularity inferences. He also avoided the pitfalls associated with the second mode by not specifying the new center’s domain of operation based solely on the specification of the patients’ language comprehension deficits.

A review of the literature will reveal that all three modes of the modularity inferences have been, and continue to be, used by cognitive scientists. Moreover, preferences for the use of one particular mode over the others seem to vary widely across disciplines, as well as among theorists within any given discipline. The question one might ask, then, is what are the reasons which make the use of one mode appear appropriate in the eyes of a theorist?

One of the main factors appears to be the theorist’s particular commitment to, and interpretation of, the functional or anatomical modularity assumption. This, in turn, might depend on the theorist’s own understanding of the basic functional and organizational principles of the central nervous system. Perhaps the best example of this lies in the sharp contrast noted earlier between Wernicke and some of his contemporaries with respect to their handling of dissociation data. While Bouillaud and his followers (including Broca), who had explicitly endorsed
Gall’s principle of radical localization of higher faculties, opted for the first mode, Wernicke clearly adopted the third mode. In this case, there can be no doubt that Wernicke’s knowledge of Meynert’s theory of cortical function played a crucial role in his theorizing. One can in fact argue that it is this particular knowledge of brain anatomy and physiology which cued him as to which cognitive processes were, and which were not likely to be localizable.

To be sure, many other factors are likely to contribute to a theorist’s decision to privilege one particular mode over the others, and there is probably no point in trying to come up with a list of criteria (or rather three such lists) that could serve as the general guideline for making such choices. In the ideal cases, however, adequate justification for the use of a particular mode would have to be made explicit by the theorist.

In Part II of the dissertation, I will explore the value and implications of adopting a particular approach to the specification of cognitive components that incorporates the third operating mode of the modularity inferences as one of its basic tenets.

II. An ambiguity in the notion of cognitive module

The working/use distinction and the above discussion of the first two operating modes of the modularity inferences suggest an ambiguity in the notion of cognitive module. On the one hand, when a modularity inference operates
according to the first mode, what is specified is the domain-specific working of
the hypothesized cognitive module. For example, as noted in section I, De Hann
(2001) uses this mode to infer that faces are processed by a separate cognitive
system dedicated to faces. Let us refer to this first kind of module as \textit{w-module}.

On the other hand, when a modularity inference operates according to the
second mode, what is specified is the domain-specific use of the hypothesized
cognitive module. For example, also noted in section I, Kanwisher et al. (1997)
use this mode to infer that the FFA constitutes a “module specialized for face
perception”. Notice here that this claim is not that face perception is the \textit{working}
of the FFA (as in the first mode), but that face perception is the restricted use of
that structure. Let us refer to this second kind of module as \textit{u-modules}.

The expression “face recognition module” is thus ambiguous in this way. Both
types of face recognition module have been hypothesized, but it is perhaps no
surprise that the \textit{w-face-recognition module} hypothesis normally occurs in the
context of the functional modularity inference. The reason for this is that high-
level cognitive operations such as face-recognition—speech production,
word/sentence recognition, mind reading, numerical cognition—which typically
figure in modularity claims, do not appear to be localizable in anatomically
distinct, and relatively small, regions of the brain. Such cognitive operations are
not, therefore, good candidates for the localization of \textit{w-modules}.

The \textit{u-face-recognition module} hypothesis, by contrast, can easily occur in the
context of the anatomical modularity inference (in addition to the functional
modularity inference), as it is not committed to the localization of a high-level
operation, but only to the localization of some unspecified part of that operation
(more on this in Chapter 4, section II).
PART TWO

The Working Zone Approach and Integration in Cognitive Neuroscience
Cognitive science is widely committed to viewing the brain as what Norman Geschwind describes as ‘more or less specialized groups of cells connected by relatively discrete pathways’ (Geschwind, 1965, p. 239; Schmahmann, & Pandya, 2006). Specialization within the brain, and therefore within the mind, is seen as an undisputed fact by virtually all cognitive scientists. What is at issue is how best to characterize this specialization. On this question, cognitive scientists’ positions have varied greatly: their views range from the early connectionist picture of Carl Wernicke to the highly modular model inspired by Jerry Fodor. As we have seen in Part I, Wernicke viewed the cortex as a collection of sensory and motor ‘images’ organized in a ‘mosaic-like arrangement’:

any higher psychic process, exceeding these mere primary assumptions, could not... be localized, but rested on the mutual interaction of these fundamental psychic elements mediated by means of their manifold connections via the association fibers (Wernicke, 1885),

At the other end of the spectrum, the functional and anatomical modularity assumptions make the mind a configuration of cognitive modules (i.e. domain specific and relatively autonomous cognitive mechanisms) realized in some
specific and relatively small regions of the brain.

In practice, the choice of experimental designs, the interpretation of experimental results, and the building of theoretical models, can all be influenced by the assumptions cognitive scientists make regarding the general organization of the mind-brain. In Chapter 1, we saw that the modular approach to the study of cognition consists of 1- a particular view of what cognitive modules are, and the extent to which they populate the mind/brain (the modularity assumptions), and 2- a method for inferring cognitive and anatomical modularity based on behavioral, and/or behaviorally derived, data (the modularity inferences).

In Chapters 2 and 3, I gave an analysis of both the anatomical and functional modularity inferences as they are currently used in various subfields of cognitive science. The two principal upshots of my analysis are 1- that the modularity inferences require the careful transition from the specification of cognitive uses to the specification of cognitive workings (Chapter 2, section I), and 2- that they can operate in three distinct modes (Chapter 3, section I).

These results, in turn, allowed for the identification of a major limitation of these inferences. This limitation is that the modularity inferences lose much of their appeal when modules are specified at too high a level.

In Part II of the dissertation, I focus on the problem of mapping cognitive functions onto the brain. More specifically, my goal is to show how the modular approach can be modified in order more successfully to decompose cognitive functions into localizable computational operations. My aim is to propose a workable alternative to the anatomical modularity assumption and to incorporate
it into a more promising approach to the mapping of cognitive functions onto the brain. According to this approach, which I call the working zone approach, the ‘function’ of a distinct brain region is specified in terms of the domain-neutral, contribution it makes to the cognitive operations in which it participates.

In this chapter, I first introduce the concept of cognitive working zone (section I), and then develop a taxonomy of four different kinds functional specifications to help expose the contrast between the working zone approach and the modular approach (section II). In section III, I use the case of the contribution of the frontal lobes to working memory operations to show how the two approaches differ in practice. In Chapter 5, I demonstrate the value and implications of adopting this approach in cognitive neuroscience. Finally, in Chapter 6, I make explicit the approach’s commitments to evolutionary thinking, and I contrast these with the adaptationist program which is currently en vogue in evolutionary psychology.

I. Cognitive working zones

A cognitive zone is any bounded region of the brain to which a cognitive ‘function’ can be attributed. Such brain regions can vary widely in size and shape depending on the particular nature and purpose of the proposed functional specification. Importantly, cognitive zones may or may not correspond to recognized anatomical structures (e.g. Brodmann areas, gyri, nuclei). For instance, each such zone may correspond to an individual or to multiple Brodmann areas (or parts thereof), or it may comprise distinct, but closely
located, sub-regions connected by identifiable fiber pathways. A cognitive working of a cognitive zone is, as we saw in Chapter 2, a cognitive operation, or set of such operations, that the structure performs. The cognitive working of a cognitive zone is, therefore, entirely contained within its anatomical boundaries, and it can be thought of as the cognitive aspect of the structure’s internal activity. (This is not to say, of course, that cognitive zones do anything much on their own—they require inputs from other parts of the brain, and also feed data outwards.) What I will call a cognitive working zone is an anatomico-functional complex—or form-working complex—the working of which is specified in a way that best captures its basic (i.e. domain-neutral) contribution to the cognitive operations in which it participates. As such, the concept of cognitive working zone denotes a particularly stable relationship between form and ‘function’ (it will become clear shortly how this kind of form-‘function’ relationship differs from other ones).

A similar kind of form-‘function’ relationship has been discussed in the biological sciences, and more specifically in the study of functional anatomy (Woodger, 1929, ch. 7; Bock & Wahlert, 1965). To use a classic example, the functional specification of the heart as a pump combines the relevant aspects of the heart’s morphology (its cavities and valves) with a particular description of its working (the performance of rhythmic contractions) into a stable anatomico-functional complex, or whole functional unit. The latter, in turn, is what best captures the heart’s basic contribution to the physiological processes in which it participates, e.g. the circulation of blood.
This way of conceptualizing the relationship between anatomical form and cognitive ‘function’ stands in sharp contrast to another way of thinking about this relationship, one in which the ‘function’ of a brain region is specified in terms of a cognitive operation, or set of such operations, in which the region is thought to participate. As we have seen, when the ‘function’ of a brain region is specified in this way, what is specified is the region’s cognitive use, and this use can be thought of as the cognitive aspect of the region’s external activity. In contrast to cognitive working zones, the relationships between brain regions and their cognitive uses cannot be thought of as stable anatomico-functional complexes. For most brain regions are likely to have different sets of cognitive uses in different cognitive contexts. Cognitive working zones, by contrast, remain invariant across very different cognitive contexts.

Broca’s area (defined here as Brodmann’s area [BA] 44) provides a good example to help make sense of these differences between cognitive working zones on the one hand, and the relationships between brain structures and their cognitive uses on the other. A growing body of evidence shows that this area is involved, together with other brain regions, in a relatively large number of linguistic and non-linguistic operations (e.g. syntactic operations in both music and natural languages, object manipulation, action perception)—i.e. that this area has many different cognitive uses in multiple informational domains (Patel, 2003; Schubotz & Fiebach, 2006). In an effort to explain these findings, cognitive neuroscientists have recently proposed a variety of cognitive operations that Broca’s area might perform that could account for its common involvement in
these various linguistic and non-linguistic operations. For example, Fiebach and Schubotz (2006) propose a model that could account for the basic contribution of Broca’s area (and the adjacent ventral premotor cortex) to a ‘wide range of attentional, cognitive, and motor processes’ (p. 499). According to their model, Broca’s area would be a ‘hypersequential processor’, i.e. a cognitive working zone that performs the ‘detection, extraction, and/or representation of regular, rule-based patterns in temporally extended events’ (p. 501). Tettamanti and Weniger (2006) make a similar, although less detailed, proposal, as they suggest that Broca’s area may be a ‘supramodal hierarchical processor’, i.e. a cognitive working zone that can ‘process hierarchical structures in a wide range of functional domains’ (p. 491).

II. A taxonomy of functional specifications for cognitive neuroscience

The foregoing discussion of the differences between cognitive working zones and the relationships between brain regions and their cognitive uses points to a pair of important distinctions that are characteristic of the way functional specification operates in cognitive science. The first one concerns the functional aspect of functional specification, and it is captured by the distinction between working and use. Functional specification can operate either via the specification of cognitive workings or via the specification of cognitive uses. The second distinction concerns the specificity aspect of functional specification, and it is
captured by the distinction made earlier between *informational* (or *domain*) specificity and *operational* specificity. Functional specification can be informationally specific (or domain-specific), or it can be operationally specific (or domain-neutral). Fig. 1 below outlines the four distinct modes of functional specification which derive from this pair of distinctions.

As can be seen from this figure, a primary difference between a cognitive working zone and a *form-working relationship* is that the latter is specified in terms of a particular informational domain, while the former is specified solely in
terms of the type of operations performed. Similarly, the distinction between informational and operational specificity allows for the identification of two additional form-'function' relationships, namely, *form-use* and *form-operation relationships*.

A more subtle difference between cognitive working zones and form-working relationships is that the latter will tend to involve cognitive operations that are more specific (or less general) than the former, even when both kinds of specification concern the same brain region. This is because form-working relationships, but not cognitive working zones, are specified in terms of particular informational domains. For instance, in the two examples given in Fig. 1, 'syntactic operations' may be seen as a specific application of 'hypersequencing operations'.

Note also that specifying either form-working or form-use relationships does not amount to claiming domain specificity, though it is compatible with there being a domain-specific application. For example, functionally specifying Broca’s area’s working as the performance of *linguistic* syntactic operation O in one context may be compatible with specifying its working as the performance of *musical* syntactic operation O’ in a different context. Similarly, pointing out the involvement of the dorsolateral PFC in *spatial* working memory in one context may be compatible with pointing out its involvement in *verbal* working memory in a different context. Claims of the existence of cognitive *modules*9 often

---

9 A reminder from Chapter 1: of all the properties usually attributed to modules, domain specificity and informational encapsulation are probably the only ones for which there is a reasonable agreement among cognitive scientists. Note, however, that there is a lack of agreement as to what level of encapsulation is required. For these reasons, I am following
accompany the specification of form-working or form-use relationships, but these claims require additional evidence and/or assumptions (Shallice, 1988).

A related point is that the specification of either cognitive working zones or form-operation relationships is not necessarily opposed to the specification of cognitive modules. For instance, functionally specifying Broca’s area as a hypersequential processor may be compatible with specifying it also as a cognitive module specialized for linguistic syntax. Similarly, pointing out the involvement of the mid-dorsolateral PFC in working memory may be compatible with specifying it as a cognitive module specialized for spatial working memory.

But there is a more important difference between the way cognitive working zones and cognitive modules are typically specified, and this is due to these two theoretical efforts having different scopes. On the one hand, as noted in Chapter 3, section II, cognitive modules are typically specified in terms of psychologically available operations—i.e. operations that ensue in psychologically available outputs, such as word/sentence recognition, speech production, face recognition, music perception. On the other hand, cognitive working zones will typically (if not almost always) be specified in terms of psychologically unavailable operations. This is because cognitive working zones, but not cognitive modules, must be specified in terms of cognitive operations that are performed by anatomically distinct, and generally relatively small, areas of the brain (see Chapter 3, section II).

For example, to say, along with Fiebach and Schubotz (2006), that Broca’s

---

Coltheart’s (1999) suggestion to use the term ‘cognitive module’ to mean a domain-specific cognitive mechanism.
area is a hypersequential processor is to say that it performs the detection, extraction, and/or representation of regular, rule-based patterns in temporally extended events (psychologically unavailable operations), but to say that the FFA is a face recognition module does not necessarily (or in fact plausibly) mean that it performs face recognition (a psychologically available operation)\(^\text{10}\). The FFA could, however, be specified as a cognitive working zone. That is, its cognitive working could be specified in terms of the cognitive operations (presumably psychologically unavailable) that best capture its domain-neutral contribution to face recognition. This could in turn help explain the FFA’s contribution to the other cognitive operations in which it appears to participate, namely, the performance of fine-grained, within-category distinctions between grossly similar visual representations of non-face objects (Gauthier, Tarr, Anderson, Skudlarski, & Gore, 1999; Tarr & Gauthier, 2000) (note, however, that a successful explanation of the FFA’s contribution to these non-face-specific operations would definitely put to rest the idea that this region is a face-recognition module, i.e. a face-specific subsystem.)

**III. Case study 3: the lateral prefrontal cortex and working memory**

Having now introduced the notion of cognitive working zone and contrasted it with other forms of functional specifications, and with modular forms in particular,
I now would like to introduce a concrete example from the cognitive neuroscience of working memory to highlight some of the important differences between the working zone approach and current uses of the modular approach.

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, such cognitive operations 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 (for a review, see Petrides, 1994). In these tasks, the animal is first presented with two identical stimuli (e.g. cups). Then, in full view of the animal, the experimenter hides a piece of food under one of the two stimuli. The stimuli are then hidden from view for a short delay period (a few seconds) by the lowering of an opaque screen. After the delay, the screen is removed and the animal is allowed to choose one of the stimuli and obtain the food if the correct stimulus is chosen. 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 operations similarly depend on the normal functioning of the lateral PFC (see Petrides, 1989 for a review).

The 1990s saw the development of two divergent theoretical frameworks regarding the involvement of the lateral PFC in working memory operations.
According to one framework, (Goldman-Rakic, 1987; 1995), the lateral PFC is organized into discrete areas, each specialized for the temporary storage of a specific type of information. This domain-specific, or modular, approach has led to the proposal that the dorsolateral PFC is an area specialized for spatial working memory while the ventrolateral PFC is specialized for object and non-spatial feature working memory. This model is often presented as a frontal extension of the ‘where’ (occipitoparietal) and ‘what’ (occipitotemporal) pathways of the primate visual system (Ungerleider, & Mishkin, 1982). Notice also that this model of PFC function primarily consists of the specification of form-use relationships. Let us call this model the domain-specific model.

An alternative framework has been proposed by Petrides (1994; 1996). According to this framework, the lateral PFC contains discrete and specialized areas that are organized according to the nature of the particular contribution that each makes to mnemonic operations, and not according to particular informational domains. This operation-specific approach has led to the proposal that the lateral PFC contains two levels of executive operations. The mid-ventrolateral PFC (BA 47/12 and 45)—in interaction with posterior cortical association areas where perception, long-term storage and transient maintenance of information occurs—mediates basic executive operations, such as the active retrieval, selection, comparison, and judgement of stimuli held in short- and long-term memory. In contrast, the mid-dorsolateral PFC (BA 9, 9/46, 46)—in interaction with the mid-ventrolateral PFC system—mediates more complex mnemonic operations, such as the monitoring and manipulation of
information within working memory. As we will see below, this model of PFC function primarily consists of the specification of cognitive working zones. Let us call this model the two-level model.

Experimental evidence for the domain-specific model has come from three research domains: electrophysiological recording experiments in monkeys; lesion studies in monkeys; and functional neuroimaging studies in humans. In electrophysiological recording experiments, the activity of single neurons in either dorsolateral or ventrolateral PFC is recorded during the delay period of spatial and non-spatial delayed-response tasks. In this way, neuronal activity in various cortical regions can be associated with the temporary maintenance of different types of information. In one set of experiments, recordings were made from the dorsolateral PFC while monkeys performed a spatial delayed-response task. It was found that a large proportion of neurons exhibiting selective neuronal activity were active during the delay period of the task (Funahashi, Bruce, & Goldman-Rakic, 1989; Funahashi, Bruce, & Goldman-Rakic, 1990). In a subsequent experiment, recordings were made from neurons in the ventrolateral PFC while monkeys performed delayed-response tasks involving either spatial stimuli (directional cues) or non-spatial stimuli (visual patterns). It was found that a greater proportion of neurons exhibiting selective neuronal activity were active during the delay period of the non-spatial delayed-response task than during the delay period of the spatial delayed-response task (Wilson, Scalaidhe, & Goldman-Rakic, 1993).

Corroboration of these findings can be found in lesion studies of monkeys.
Thus, impairment on spatial working memory tasks has been observed in monkeys with lesions to the dorsolateral PFC (Funahashi, Bruce, & Goldman-Rakic, 1993), whereas impairment on non-spatial working memory tasks has been observed in monkeys with lesions to the ventrolateral PFC (Passingham, 1975; Mishkin & Manning, 1978).

Direct evidence in support of the domain-specific model in humans has come from functional neuroimaging experiments. Thus, portions of the dorsolateral PFC have been found to show greater activation as subjects engage in visuo-spatial working memory tasks (e.g. McCarthy et al., 1994; Courtney, Ungerleider, Keil, & Haxby, 1996; Courtney, Petit, Maisog, Ungerleider, & Haxby, 1998). In contrast, portions of the ventrolateral PFC have been found to show greater activation as subjects engage in non-spatial visual working memory tasks (e.g. Cohen et al., 1994; Courtney, Ungerleider, Keil, & Haxby, 1996; Courtney, Petit, Maisog, Ungerleider, & Haxby, 1998).

Taken together, these electrophysiological, lesion, and neuroimaging findings in both monkeys and humans have been interpreted as evidence of the modular organization of working memory operations in the lateral PFC (Goldman-Rakic, 1995; 1996). It is also important to notice that all these lines of evidence exemplify rather nicely what I have described as the second mode of the anatomical modularity inference (Chapter 3, section I), which consists of specifying the domain of operation of hypothesized components (here spatial information for the dorsolateral PFC and non-spatial information for the ventrolateral PFC) without specifying their respective working, or by specifying
only their cognitive use.

The development of the two-level model has been part of an effort to capture the basic contribution of the lateral PFC to mnemonic operations. The introduction of a novel working memory task—the self-ordered task—by Petrides and Milner (1982) served as a starting point for this line of research. 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. Petrides (1994) has referred to this capacity as the ‘monitoring of events within working memory’.

When tested on four self-ordered tasks—two verbal and two nonverbal—it was found that patients with lesions limited to the lateral PFC, but not patients with lesions limited to the temporal lobes, were severely impaired on all four tasks, although they could perform well on several other short-term memory tasks (Petrides & Milner, 1982). Subsequent work in monkeys, where lesions can be restricted to precise locations, has revealed that lesions limited to the mid-dorsolateral PFC are sufficient to severely impair performance on the self-ordered task (Petrides, 1991; 1995). Such lesions to the mid-dorsolateral PFC, however, have been found not to impair performance on various delayed-response tasks (Mishkin, Vest, Waxler, & Rosvold, 1969; Passingham, 1975).
These findings, when combined with earlier observations in monkeys that lesions to the ventrolateral PFC result in severe impairments on both spatial and non-spatial delayed-response tasks (Mishkin, Vest, Waxler, & Rosvold, 1969; Mishkin & Manning, 1978) strongly suggest that there is a fundamental difference in the particular contribution made by different parts of the lateral PFC to working memory operations (Petrides, 1994; 1996). Notice here that this kind of reasoning exemplifies the third mode of the anatomical modularity inference, where a dissociation between two cognitive tasks (here the self-ordered and delayed response tasks) is used to infer the existence of distinct cognitive components with restricted, yet unspecified, domains of operation. This is in contrast to the reasoning behind the domain-specific model which we said exemplifies the second mode, as noted above.

This fundamental difference in the contribution of the mid-dorsolateral and mid-ventrolateral PFC to working memory processes is represented in the two-level model by two cognitive working zones corresponding to the two levels of executive control within this area of the brain. The first zone is the mid-ventrolateral PFC where the first level of processing occurs. This processing consists in the initiation of active retrieval, and active encoding, of information held in short- and long-term memory. These operations, in interaction with posterior cortical association areas where long-term storage and transient maintenance of information occurs, are crucial for the performance of various first-order executive operations such as the active retrieval, judgement, selection, and comparison of mnemonic information. The second zone is the mid-
dorsolateral PFC where the second level of processing occurs. This processing consists of the recoding in abstract form, and online maintenance, of multiple pieces of information held in short- and long-term memory. These operations, in interaction with the mid-ventrolateral PFC, are crucial for the performance of more complex executive operations such as the monitoring and manipulation of information within working memory.

More recently, support for the two-level model has come 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. One of these tasks, the self-ordered task, has been described above. A variation of this task is the externally ordered task, in which the subject must keep track of the (random) presentation of stimuli from a known set, and is asked to identify the stimulus that has been left out in the presentation. In a third task—the n-back task—the subject observes a series of stimuli on a visual display and must respond according to whether or not the current stimulus matches the stimulus that was presented a specified number of steps earlier (e.g. two steps back). According to the two-level model, all of these tasks should engage the mid-dorsolateral PFC, as they all require the online maintenance of multiple stimuli for monitoring purposes. Using either positron emission tomography (PET) or functional magnetic resonance imaging (fMRI), normal human subjects were scanned as they performed a non-spatial visual self-ordered task (Petrides, Alivisatos, Evans, & Meyer, 1993), verbal self-ordered and externally ordered tasks (Petrides, Alivisatos, Meyer, &
Evans, 1993), 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.

Functional neuroimaging experiments have also been conducted that show a systematic relationship between the activation of the mid-ventrolateral PFC and a requirement for active retrieval of information held in short- and long-term memory. According to the two-level model, active retrieval must be distinguished from automatic stimulus-driven or context-driven retrieval of mnemonic information. In the latter, stimulus retrieval is driven by strong, stable and unambiguous posterior cortical associations between stimuli, or between stimuli and their contexts, and it is therefore independent of the frontal lobes\textsuperscript{11}. Active retrieval, in contrast, requires the top-down modulation, by the mid-ventrolateral PFC, of memory traces held in posterior cortical association areas when no such stable, strong, or unambiguous relations exist between particular memory traces, contexts, and responses (Petrides, 2002). One prediction of the model, therefore, is that the mid-ventrolateral PFC will be recruited by any task that requires the active retrieval of information held in short- and long-term memory. This prediction was tested in studies with normal subjects performing verbal (Petrides, Alivisatos, & Evans, 1995), spatial (Owen, Evans, & Petrides, 1996), non-spatial (Petrides, Alivisatos, & Frey, 2002), and combined spatial/non-spatial (Cadoret,

\textsuperscript{11} This hypothesis is supported by 1- the fact that performance on several recognition memory tasks can be normal after lateral frontal lesions, and 2- functional neuroimaging data on the differential activation of the mid-ventrolateral PFC during highly automatized versus highly explicit retrieval conditions. See Petrides (1996) for details.
Pike, & Petrides, 2001) 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.

Finally, the two-level model makes two additional predictions in the form of a double dissociation between mid-ventrolateral and mid-dorsolateral cognitive processing within the PFC. According to the first prediction (P1), the mid-ventrolateral, but not the mid-dorsolateral PFC, will be engaged by a task’s requirement for active retrieval of mnemonic information. According to the second prediction (P2), the mid-dorsolateral, but not the mid-ventrolateral PFC, will be recruited by a task’s additional requirement for online maintenance of multiple pieces of information for monitoring and manipulation purposes. In order to test these predictions, Petrides et al. (2002) have conducted a PET experiment in which normal subjects were scanned while performing three different tasks. In the control task, subjects were presented with pairs of colored abstract designs on a touch-sensitive screen, and were instructed to look at the designs for a moment and then touch the screen between the two to move to the next pair. Half of the designs had been seen once by the subjects just before the experiment, the other ones were completely novel. In the familiarity/novelty task, subjects were presented with similar pairs of abstract designs, one of which had been seen just before the scanning, and had to select the novel one by touching it in order to move to the other pair. Thus, the only difference between this and the control task is that the former required the subjects to make an explicit
judgement on the stimuli, by actively encoding them as novel or new, as opposed to passively viewing novel and familiar stimuli. When compared to activation during the control task (where no active encoding is required) activation during the familiarity/novelty task was significantly greater in the mid-ventrolateral PFC, but not in the mid-dorsolateral PFC, in agreement with P1. In the monitoring task, the subjects were again presented with pairs of abstract designs and had to select one of the two designs in order to advance to the next pair. They were also told that some of the pairs presented would recur, in which case they would have to select the design that they had not previously selected (to make the proportions of novel and familiar stimuli the same as with the other tasks, half of the pairs were repetitions). As in the familiarity/novelty tasks, this task required the subjects to make active familiarity judgments. Unlike the familiarity/novelty task, however, successful performance on the monitoring task additionally required the subjects to keep track of (i.e. monitor) their previous choices. When compared to activation during the control task, activation during the monitoring task was greater in both the mid-ventrolateral and mid-dorsolateral PFC, as predicted by the model. Moreover, when compared to activation during the familiarity/novelty task, activation during the monitoring task was greater in the mid-dorsolateral, but not the mid-ventrolateral PFC, in agreement with P2.

On the surface, the domain-specific and the two-level models can be said to differ in the particular functions ascribed to various lateral prefrontal cortical regions (Owen, 1997). However, the use of the term ‘function’ here easily hides
some important differences in theoretical approach between these two efforts. These differences are best captured by considering the kind of functional specification that each model is trying to establish. On the one hand, the domain-specific model specifies form-use relationships that are hypothesized on the basis of the second mode of the anatomical modularity inference. Thus, the model attributes different domain-specific cognitive uses to the ventrolateral and dorsolateral PFC; the latter is specialized for spatial working memory while the former is specialized for non-spatial (or object) working memory. On the other hand, the two-level model specifies cognitive working zones that are hypothesized partly on the basis of the third mode of the anatomical modularity inference. Two such zones are described by the model. One refers to the mid-ventrolateral PFC as an area in which active retrieval initiation and encoding operations take place, and the other refers to the mid-dorsolateral PFC as an area in which abstract recoding and online maintenance operations occur. Both cognitive working zones are then used to account for some of the cognitive uses that each of these two working zones have in working memory operations.

In the next chapter, I give an analysis of the value and implications of adopting this working zone approach, as exemplified by the development of the two-level model, for the study of how cognition is implemented in the brain.
It is a primary goal of cognitive neuroscience to map cognitive functions onto the brain. This task is highly sensitive to methodology, because it varies greatly with how cognitive functions are specified. But there is no generally agreed method of functional specification in the diverse disciplines within this field. This situation poses a serious challenge to another widely acknowledged goal of cognitive neuroscientific research, namely, the coordination of research efforts within particular disciplines and across disciplinary boundaries through the integration of theories, methods, and assumptions. One aspect of this integration challenge is that the functional specification of brain regions can, as we have seen, operate in two importantly different modes, which correspond to two different uses of the term ‘function’. In the first mode, the ‘function’ of a particular brain region is specified in terms of a cognitive operation, or set of such operations, that this region is thought to perform (i.e. carry out internally). In this case, what is specified is the region’s cognitive working. In the second mode, the ‘function’ of a particular brain region is specified in terms of a cognitive operation, or set of such operations, in which this region is thought to take part, or participate. In this case, what is specified is the region's cognitive use.

The distinction between these two modes of functional specification often goes unnoticed in much cognitive theorizing—notice how easy it is to miss the
distinction between the two modes if one says both 1- that the abstract recoding and online maintenance of information is a *function* of the mid-dorsolateral PFC (here ‘function’ refers to cognitive operations that are carried out internally by the mid-dorsolateral PFC), and 2- that face recognition is a *function* of the FFA (here ‘function’ refers to a cognitive operation that goes far beyond FFA’s boundaries). However, as we will see, these two modes can have importantly different effects on the development of cognitive models and theories.

A second aspect of the integration challenge is that each of these two modes of functional specification can, as we have also seen, take two different forms. On the one hand, functional specification can be called *informationally specific* (or *domain-specific*) when it is done in terms of stimuli, or inputs, of a particular class, e.g. faces, colors, numbers. An example of this is the claim that the dorsolateral PFC is specialized for spatial working memory. On the other hand, functional specification can be called *operation specific* (or *domain-neutral*) when it is done in terms of operations that are in principle applicable to any class of stimuli. An example of this is the claim that the mid-dorsolateral prefrontal cortex PFC performs the abstract recoding and online maintenance of multiple pieces of information.

This chapter tackles both of these aspects of the integration challenge. The main goal is to give an account of the value and implications of 1- specifying brain regions in terms of cognitive workings (as opposed to cognitive uses), and 2- specifying these cognitive workings in terms of domain-neutral operation types. The payoff of specifying ‘functions’ in this way is, I will argue, a potential
for integration in cognitive neuroscience. Three integrative aspects are discussed. The first is general and it concerns the integration of models and theories within and across disciplines. The second concerns the use of functional neuroimaging data to distinguish between competing cognitive theories. And the third concerns the building of animal models of human cognition.

I. Horizontal and vertical integration

One significant integrative aspect of the working zone approach is that the specification of a cognitive working zone will often have the potential to account for the basic contribution of a given brain structure to a wide range of (sometimes seemingly unrelated) cognitive operations. We may call this horizontal integration¹². One way that this kind of integration can be accomplished is by combining two or more independent lines of research pertaining to different informational domains. This was illustrated in section 1 with the example of Broca’s area and its involvement in a multitude of linguistic and non-linguistic operations. The suggestion that this area might work as a hypersequential processor (Fiebach & Schubotz, 2006), or as a supramodal hierarchical processor (Tettamanti & Weniger, 2006), is a direct result of efforts to combine a wide range of findings in attentional, cognitive, and sensorimotor research. Tettamanti and Weniger (2006) describe their approach as follows:

¹² A related point is that an adequate account of the basic cognitive working of a given brain region may require the consideration of its participation across multiple task categories (Anderson, 2007a).
a perspective that has not been explored in great detail is that of isolating the structural properties of tasks involving activity in Broca’s area and examining them with respect to a cognitive theory, giving particular attention to the computational demands involved (p. 491)

Thus, while the specification of form-use relationships within several lines of research on Broca’s area have so far yielded a set of disparate domain-specific cognitive components (or modules) (Schubotz & Fiebach, 2006), the approach which consists in specifying cognitive working zones emerges as a promising alternative.

The above passage also suggests another important integrative aspect of the working zone approach. Notice first that there are two general approaches to specifying cognitive operations, based on whether or not cognitive theorizing is concerned with the localization and/or implementation of cognitive operations in the brain. On the one hand, functional specification can be called brain-based if it is concerned with the localization and/or implementation of cognitive operations in the brain. On the other hand, functional specification can be called non-brain-based if it is not concerned with the localization and/or implementation of cognitive operations in the brain. In neuropsychological research—the making of inferences about the architecture of the normal mind-brain based on the behavioral effects of brain lesions—the split between these two approaches took place more than a hundred years ago with the emergence of neurolinguistics in
the second half of the nineteenth century. While every element of Wernicke’s (1874) model of language processing corresponded to well defined structures of the brain (anatomical centers and fiber tracts), many subsequent models, including Lichtheim’s (1885) elaboration of Wernicke’s model, incorporated several centers and pathways that had no known anatomical correlates (see Morton, 1984 for a discussion of this point).

Today, the split between these two approaches is even more pronounced, as some subfields of cognitive science are being concerned mainly with one of the two approaches. For example, workers in the field of cognitive neuroscience primarily adopt the brain-based approach to functional specification. An example of this is Petrides (1996, 2005) model of the involvement of the lateral PFC in working memory operations (Chapter 4, section II). By contrast, workers in the field of cognitive psychology primarily adopt the non-brain-based approach to functional specification. An example of this is Coltheart et al. (2001) computational model of visual word recognition and reading aloud.

Now, as the above example of Broca’s area illustrates, the specification of cognitive working zones requires the consideration of multiple levels of analysis ranging from the cognitive level, both representational and computational, to the anatomical and neurophysiological levels—i.e. it requires a methodology that cuts across the brain-based/non-brain-based divide. We may call this vertical integration.

Another illustration of these two integrative aspects of the working zone approach is the case of the parietal cortex as a site of interaction between
numerical and spatial cognition. A clear demonstration of the interaction between numbers and space is the spatial-numerical association of response codes (SNARC) effect (Dehaene, Bossini, & Giraux, 1993). The SNARC effect is easily demonstrated by asking subjects to classify numbers as even or odd by making their responses on either the right or the left side of space. What is found is that responses to larger numbers are faster when made on the right side of space whereas responses to smaller numbers are faster when made on the left side of space. In their review of the literature, Hubbard et al. (2005) note that the SNARC effect can also arise in non-numerical tasks (e.g. judging the phonemic content of number words), and that such spatial-numerical interactions seem to occur at a level of processing that is both task- and modality-independent. In addition to reviewing the behavioral evidence in humans, Hubbard and colleagues also discuss the findings from two other lines of research which, until recently, had been pursued largely independently from one another. In monkeys, several neurophysiological studies point to a set of parietal regions involved in various aspects of spatial processing. In humans, neuroimaging studies have repeatedly associated the abstract representation of quantity with activity in the intraparietal sulcus. On the basis of these findings, as well as more recent studies that show a significant overlap in the parietal basis of numerical and spatial processing in both species, the group proposes that the observed interactions between numerical and spatial cognition ‘arise from common parietal circuits for attention to external space and internal representations of numbers’ (Hubbard, Piazza, Pinel, & Dehaene, 2005, p. 435). What they propose, in fact, is
a cognitive working zone composed of the lateral and ventral regions of the intraparietal sulcus as well as the fibers connecting the two regions, and in which ‘the same computational transformations that support spatial updating are crucial for arithmetic operations that create shifts in the locus of activation along an internal number line’ (p. 445).

In this example, the specification of a cognitive working zone accounts for the basic contribution of a brain region to cognitive operations pertaining to two different informational domains, namely, numbers and space (horizontal integration). It also derives from the consideration of several different methodologies at multiple levels of analysis (vertical integration).

II. Functional neuroimaging and cognitive theory

Functional neuroimaging—restricted here to functional Magnetic Resonance Imaging (fMRI) and Positron Emission Tomography (PET)—is a widely used technique for studying the organization of cognitive operations in the human brain. It is by far the preferred tool for localizing cognitive operations in specific brain regions. It is less clear, however, to what extent functional neuroimaging can be used effectively to study cognition itself. That is, it is not clear to what extent this method of investigation can help advance cognitive theory. On the one hand, some have argued that functional neuroimaging in principle is seriously limited in this domain (Van Orden & Paap, 1997; Uttal, 2001; Page,
2006). On the other hand, there now seems to be an industry of claims reporting the successes of such attempts (see e.g. Henson, 2005 for a recent overview of these efforts in the context of memory research, as well as the articles in a recent forum on this topic published in the April 2006 issue of Cortex). The proposed framework can help give a clearer picture of how functional neuroimaging can serve as an effective tool for the study of cognition by allowing for a more refined analysis of one of the main limitations of this technique when it is used to that effect.

One of the principal ways that functional neuroimaging data can inform cognitive theorizing is to use such data to distinguish between competing cognitive theories, i.e. theories about some aspect of cognition that are expressed at the psychological level, or, to use the terminology introduced earlier, non-brain-based theories. Let us begin with an example.

Jack et al. (2006) describe two competing cognitive theories of inefficient visual search—a form of visual search in which the time required to identify the target is proportional to the size of the set to be searched. According to the first theory (T\textsubscript{a}), ‘inefficient visual search uses a parallel search mechanism’. According to the second theory (T\textsubscript{b}), ‘inefficient visual search uses a serial search mechanism’ (Jack, Sylvester, & Corbetta, 2006, p. 420). The authors reason as follows. On the one hand, if inefficient visual search uses a parallel mechanism, then the searching process should not involve any covert shift in the location of attention. If, on the other hand, inefficient visual search proceeds serially, then the searching process should be characterized by the presence of multiple covert
attentional shifts to discrete locations. Thus, the two theories yield two incompatible predictions; either inefficient visual search involves covert shifts of attention, in which case $T_b$ must be true, or it does not, in which case $T_a$ must be true.

Using the same reasoning, and in order to distinguish between these two theories, Corbetta et al. (1995) conducted a PET experiment in which normal subjects were scanned while performing both efficient and inefficient visual searches (efficient search was used here as a control task since both $T_a$ and $T_b$ predict that it involves parallel search mechanisms). Performance on the inefficient search task relative to the control task resulted in greater activity in the posterior parietal cortex, with the strongest activation in the right posterior parietal cortex. This finding, combined with previous neuroimaging data that show that the same region in the right posterior parietal cortex is involved in the task of covertly shifting visual attention (Corbetta, Miezin, Shulman, & Petersen, 1993), was interpreted by Jack et al. (2006) as evidence in favor of $T_b$.

We can now ask: is this an example in which functional neuroimaging data have been used successfully to distinguish between competing cognitive theories? In his review of the case, Coltheart (2006) argues that it is not. He gives the following reason:

[The reasoning that is used by Jack et al. (2006)] required that covert shifting of visual attention and activation of this right posterior parietal region be co-extensive. Two things must be true: whenever there is covert shifting of visual
attention there must always be activation of this brain region, and whenever this brain region is active there must always be covert shifting of visual attention going on... But subsequent work has shown that covert shifting of visual attention is not the sole function of this parietal region (p. 426, original emphasis).

Immediately after this passage, Coltheart mentions two functional neuroimaging studies, both involving visual search tasks, in which greater activation was found in the relevant right parietal region even though the tasks in these studies did not require any covert shifting of visual attention. Certainly, this shows that greater activation in this region of right posterior parietal cortex is not always associated with covert shifting of visual attention, which means that this cognitive operation is not co-extensive with activation in this region. Therefore, Coltheart argues, this set of neuroimaging data cannot be used to distinguish between the two theories.

Jack et al. (2006) are well aware that greater activation in right posterior parietal cortex is not always associated with covert shifting of visual attention. Like Coltheart, they point to a study by Wojciulik and Kanwisher (1999) which shows that the involvement of the posterior parietal cortex in visual attention is more general than covert shifting of visual attention. They are thus equally aware that the Corbetta et al. (1995) study does not, by itself, resolve the question of whether inefficient visual search uses serial or parallel mechanisms. However, the fact that these authors point to a number of subsequent studies which lend support to Corbetta et al.’s (1995) original conclusion suggests that they would
disagree with Coltheart’s claim that their reasoning is defective *because* it does not meet the co-extension requirement.

Coltheart’s co-extension requirement (CR) can be stated as follows:

**CR**: cognitive operation \(O\) is said to be *co-extensive* with activation in brain region \(R\) if and only if,

1. whenever \(O\) is engaged, \(R\) is *always* active, and
2. whenever \(R\) is active, \(O\) is *always* engaged.

Now, consider the following inference, known as the ‘reverse neuroimaging inference’ (RNI) (Poldrack, 2006):

**RNI**: if brain region \(R\) is active, then cognitive operation \(O\) must be engaged.

Poldrack (2006) gives the following account of the reasoning behind RNI:

1. In the present study, when task comparison \(A\) was presented, brain region \(R\) was active.
2. In other studies, when cognitive operation \(O\) was putatively engaged, then brain region \(R\) was active.
(3) Therefore, the activity of region $R$ in the present study demonstrates engagement of cognitive operation $O$ by task comparison $A$ (adapted from Poldrack 2006, p. 59.)

Understood in this way, the reverse neuroimaging inference (henceforth ‘reverse inference’) is an inference to the best explanation. In this case, one concludes that task comparison $A$ must have engaged cognitive operation $O$ on the grounds that this hypothesis best explains the observed activity in brain region $R$, given the fact that the putative engagement of $O$ is known to be associated with activity in brain region $R$. The same reasoning structure is used by cognitive scientists to infer modular architectures on the basis of double dissociations. In this case, one concludes that a cognitive system must have a modular architecture $A$ on the grounds that this hypothesis best explains an observed double dissociation $D$ in that system, given the fact that such architectures are known to produce double dissociations when damaged in two different ways (Coltheart, 2001). In all such cases, the preferred explanation is contrasted against other plausible explanations. For instance, in the neuroimaging case, it is plausible that brain region $R$ be recruited by a cognitive operation other than $O$. In the neuropsychological case, we know that non-modular architectures can also produce double dissociations when damaged in two different ways (Shallice, 1988). In any given case, the less plausible these alternative explanations appear to be, the stronger the inference is.
In discussing their case, Jack et al. (2006) make use of the same reasoning. Theirs is a reverse inference in which the engagement of covert shifting of visual attention is inferred from the activation of the right posterior parietal cortex. They mention studies in which the shifting of visual attention has been associated with activity in this region. Each of these studies, they claim, adds some force to their inference. But they also mention a study in which increased activation in right posterior parietal cortex occurs in the absence of covert shifting of attention. They believe that this finding weakens their inference. Overall, they argue that the functional neuroimaging data that are currently available provide substantial support for the claim that greater activation in right posterior parietal cortex during inefficient visual search indicates the engagement of covert shifting of visual attention.

Poldrack (2006) nicely sums up the foregoing analysis. He writes: ‘the greatest determinant of the strength of a reverse inference is the degree to which the region of interest is selectively activated by the cognitive process of interest’ (p. 60). The more selectively activated a brain region is by a given cognitive operation, the stronger the reverse inference is. This invites the question: to which degree must the region of interest be selectively activated by the cognitive process of interest in order for the reverse inference to be considered strong and legitimate? There is currently no consensus on what the answer to this question should be. Notice, for instance, the size of the gap between the level of selectivity of activation prescribed by CR—i.e. total selectivity—and the level of selectivity that Jack et al. (2006) consider to be adequate. To be sure, the answer to this
question will depend, at least in part, on the nature of the case under consideration. In particular, the legitimacy of a reverse inference is likely to depend on the nature of the tasks involved, on the composition of the neuroimaging data set being used (including the size and precise location of the brain region involved), as well as on other aspects of the experimental design. However, if the use of this kind of inference is to have a significant impact on the development and evaluation of cognitive theories, some sort of consensus will have to emerge among cognitive scientists as to what a strong and legitimate reverse inference must look like.

There are two ways in which the proposed framework can help us make progress in this direction. First, it can help provide a better assessment of the main limitation associated with the reverse inference as it is currently used. Second, it can help give a clear account of the conditions under which the inference is both legitimate and strong. Let us take these in turn.

A central aspect of the working zone approach is the observation that the functional specification of cognitive components can operate in four distinct modes. Brain regions can be specified as form-use, form-operation, or form-working relationships, or as cognitive working zones. Now, the reverse inference essentially operates via the specification of form-‘function’ relationships. More specifically, it specifies implication relations between form and ‘function’. This means that it too can operate in four distinct modes corresponding to the four kinds of functional specifications just mentioned. It is therefore interesting to note that current uses of this inference almost always involve the specification of form-
use (and sometimes form-operation) relationships. This is true of all the
examples discussed by Henson (2005) and by the contributors to the previously
mentioned Cortex forum in which this inference is used. For example, Rees et al.
(1999, discussed in Henson, 2005) identify four regions of the brain that are more
activated by detecting concrete nouns than by detecting random consonant
strings; Ramnani and Miall (2004, discussed in Henson, 2005) refer to the
paracingulate cortex and the superior temporal sulcus as ‘areas most
consistently activated when subjects evaluate the intentions of others’ (p. 85);
and Seron and Fias (2006, contributors to the Cortex forum) refer to the
intraparietal sulcus as an area that is ‘activated in number semantic tasks’ (p.
407). Poldrack’s (2006) analysis of the inference also involves the specification of
form-use relationships. For example, he discusses the case of Broca’s area, and
whether activation in this area ‘implies the engagement of language function’ (p.
60).

The degree to which a given brain region is selectively activated by a
particular cognitive operation $O$ is, as we noted earlier, the greatest determinant
of the strength of a reverse inference. When the latter operates via the
specification of form-use or form-operation relationships, the selectivity of
activation of the brain region for $O$ varies according to how many different
cognitive operations the region participates in—i.e. how many different cognitive
uses it has—such that the more cognitive uses it has, the less selective its
activation is for $O$. This means that the more different cognitive uses a brain
region has, the weaker the reverse inferences involving that region will be for any
of its cognitive uses. For example, each of the brain regions mentioned in the previous paragraph are known to be recruited by several different cognitive operations pertaining to multiple informational domains—i.e. each of these regions is known to have a wide range of cognitive uses. Consequently, reverse inferences involving these regions are more likely to be seen as weak and/or illegitimate when they operate via the specification of either form-operation or form-use relationships\textsuperscript{13}.

Compare this to the case in which the reverse inference operates via the specification of cognitive working zones. Unlike form-use and form-operation relationships, cognitive working zones consist of stable anatomico-functional relationships. While the cognitive uses of brain structures tend to vary greatly according to the cognitive context in which they operate, cognitive working zones remain stable across multiple cognitive contexts. Thus, while Broca’s area appears to have very different cognitive uses depending on whether it participates in music perception, language production, or the planning and perception of actions, its contribution to this wide range of cognitive operations can be hypothesized to be the work of a single cognitive working zone (Fiebach & Schubotz, 2006; Tettamanti & Weniger, 2006). Similarly, as discussed earlier, while a region of the parietal cortex appears to be involved in a wide range of numerical and spatial operations, the specific contribution that this region makes

\textsuperscript{13} A similar argument can be made in the case of form-working relationships. In this case—i.e. when the reverse inference operates via the specification of form-working relationships—the selectivity of activation of the brain region for a particular domain-specific cognitive operation $O$ varies according to how domain-specific (or domain-general) the region is, such that the more domain-general it is, the less selective its activation is for $O$. This means that the more domain-general a brain region is, the weaker the reverse inferences involving that region are for any of its informationally specified cognitive workings.
to these cognitive operations can be understood as the work of a cognitive working zone, as specified in Hubbard et al. (2005) model of the lateral and ventral portions of the intraparietal sulcus.

Now, this greater stability in the relationship between the form and the working of cognitive working zones translates itself into the greater selectivity of activation of these zones for the cognitive operations that are specified as their workings. And since the degree to which a brain region is selectively activated by a cognitive operation is the most reliable determinant of the strength of a reverse inference, the latter will be much stronger when it operates via the specification of cognitive working zones.

The foregoing analysis provides a clear account of how functional neuroimaging data can help advance cognitive theory. The key to this is to realize that the strength of the reverse inference largely depends on which kind of functional specification it operates on. When it is used in conjunction with the specification of cognitive working zones, the inference relies on stable anatomico-functional relationships and is therefore stronger than when it is used in conjunction with the looser form-use, form-operation, and form-working relationships.
III. Evolutionary considerations and animal models of human cognition

The evolutionary approach is an important feature of the working zone approach. In many cases, the successful specification of cognitive working zones in the human brain depends on the identification of homologous regions (regions with common evolutionary origin) in the brain of other species. For example, crucial to the development of Hubbard et al. (2005) model of the human lateral and ventral portions of the intraparietal sulcus (section I) was the identification of a homologue of the model’s cognitive working zone in the brain of monkeys. In this section, I argue that for the purpose of building animal models of human cognition, homologous regions are best functionally specified as cognitive working zones, as opposed to form-use, form-operation, or form-working relationships.

The task of searching for the homologue of a human brain region in the brain of another species primarily involves the comparative anatomical and neurodevelopmental study of the two candidate regions. This task does not normally involve an extensive analysis of the regions’ comparative ‘functions’. This is so, because homologies, unlike analogies (or homoplasies), are not defined in terms of functional similarity\(^\text{14}\). In fact, 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

\(^{14}\) Two structures are said to be analogous if they possess the same or similar function without sharing a common evolutionary origin.
by common phylogenetic origin, or by sameness of position in a developmental cascade\textsuperscript{15}.

Thus, homologous structures may sometimes share very little (or nothing at all) in terms of functional similarity. However, for the purpose of building animal models of human cognition, homologous regions are interesting to the extent that they are \textit{functionally} similar—i.e. the more functionally similar homologous regions are, the more interesting they are likely to be for building animal models of human cognition. Now, the more distant a homology relation is the more likely it is that functions will be different. In accounting for the function of some organ in a homology scheme we are interested in the emergence of that function by separation from other functions of "same organs". Thus, we are interested in closely related organs that have highly comparable functions.

For example, the finding that there appears to be a homologue of Broca’s area in the ventrolateral PFC of the macaque monkey (Petrides, Cadoret, & Mackey, 2005) is interesting from a neurocognitive perspective not because it shows that this area shares a common evolutionary origin with Broca’s area, or because it is located roughly at the same place, but because it shows that it possesses some of the same important functional properties as Broca’s area (e.g. both are involved with orofacial musculature.)

One way to justify homology thinking about brain regions is, therefore, to appeal to the notion of \textit{functional} homology. The latter may at first appear

\textsuperscript{15} Both definitions are hierarchical in that two items can have a common distant origin but not a common recent origin; similarly they can have the same position in a relatively sparse account of development, but different in a more detailed account. Thus two items may belong to a single sameness class, but to different sameness sub-classes. Distance of relationship is measured in this way.
contradictory, since homology, as we have just seen, is traditionally defined as sameness of structure regardless of function. However, a group of philosophers have recently proposed plausible ways to understand the notion of homology which make room for psychological or cognitive homologies (Ereshefsky, 2007; Love, 2007; Matthen, 2000; 2007). For example, Matthen (2007) shows how different kinds of vertebrate vision (e.g. avian and mammalian vision) can be homologous despite having many different visual functions. The reason for this is that all vertebrate visual systems share a common set of core (i.e. homologous) functional components, which include homologous eye structures (e.g. lenses, retinas), homologous subcortical structures (e.g. homologues of the lateral geniculate nucleus), and homologous cortical structures (e.g. homologues of the primary and extra-striate cortices). In what follows, I argue that this kind of homology thinking about cognition can benefit from adopting the working zone approach.

To begin with, notice that the possibility of introducing homology thinking into explanations of cognitive functions invites the question of which mode of functional specification can best capture the important functional similarities between the homologous components of cognitive systems.

To answer this question, let us first notice that just as in the non-comparative case, the comparative approach to functional specification can operate via the four distinct modes previously identified. Now, the crucial point is that in the comparative case, these different modes will often give rise to very different accounts of the level of functional similarity (or dissimilarity) between two
homologous regions. To see this, let us consider what it takes, in terms of brain evolution, to effect significant changes to a region’s set of domain-specific cognitive uses (specified as form-use relationships), and compare this to what it takes to effect significant changes to a region’s basic cognitive working (specified as a cognitive working zone)\(^{16}\).

In the course of brain evolution, significant changes to a region’s set of domain-specific cognitive uses can occur for any of the following reasons:

1. Such changes can occur as a result of modifications to the way the region is directly connected to other regions (direct connectivity modifications), or as a result of modifications to the way it is indirectly connected to other regions (indirect connectivity modifications). This is so, because the cognitive uses of a brain region, as we have seen, depend on the particular interactions it has with the other brain regions with which it is connected. For example, each of Broca’s area’s hypothesized cognitive uses—speech production, music perception, object manipulation, etc.—depends on its interaction with several other regions (e.g. various sensory and motor regions) with which it is connected. Thus, any significant changes to the way this area is connected with these other regions is likely to alter the set of domain-specific cognitive operations in which it participates (i.e. it is likely to alter this area’s set of domain-specific cognitive uses.) In fact, it is by way of such direct and indirect connectivity modifications that Broca’s area may have been recruited for its new set of linguistic uses in

\(^{16}\) For the sake of brevity, I am only concerned here the comparative value of specifying form-use relationships and cognitive working zones for the purpose of building animal models of human cognition. The specification of form-working or form-operation relationships could be substituted for the specification of form-use relationships without affecting the general argument.
humans (Deacon, 1997, ch. 10).

2. Changes to a region’s set of domain-specific cognitive uses can also occur as a result of modifications to the internal organization (and thus the working) of the other brain regions with which it is directly or indirectly connected, as well as modifications to its own internal organization. For example, Broca’s area’s new set of linguistic uses in humans may have been the result of significant modifications to the internal organization of another ‘language’ area with which it is connected, namely the area traditionally known as Wernicke’s area.

3. Finally, significant changes to a region’s set of domain-specific cognitive uses can occur as a result of changes in the organism’s environment. These, in the case that concerns us, include changes that can occur in the absence of major modifications to the brain, such as changes in habitat, as well as technological, social, and cultural changes. For example, with the invention of both writing and arithmetic— inventions that are too recent to have influenced human brain evolution—several regions of the brain have acquired new domain-specific cognitive uses, as evidenced by the presence of reproducible form-use relationships between particular brain regions and these cognitive capacities (Dehaene, 2005; Dehaene & Cohen, 2007).

In contrast, significant changes to a region’s basic cognitive working can only occur as a result of modifications to its own internal organization17. This means that in the course of evolution, significant changes to a region’s set of domain-specific cognitive uses can be the result of changes in the organism’s environment and not in the organism’s basic cognitive working. This is because the basic cognitive working zone is not connected to other structures and thus does not require functional specification changes. However, such extensive changes to a cognitive working zone’s external connectivity would always co-occur with, and often result from, significant changes to its internal organization (see Striedter, 2005, ch. 7, for a discussion of this point).

17 True, extensive changes in the way a cognitive working zone is connected to other structures may require a change in the way it should be functionally specified. However, such extensive changes to a cognitive working zone’s external connectivity would always co-occur with, and often result from, significant changes to its internal organization (see Striedter, 2005, ch. 7, for a discussion of this point).
specific cognitive uses will often occur in the absence of significant changes to its basic cognitive working. This, in turn, means that when they are specified as form-use relationships, two homologous regions may more easily appear to be functionally dissimilar than when they are specified as cognitive working zones.

A consequence of this is that focusing on the specification of form-use relationships when searching for the homologue of a human brain region in the brain of another animal can give rise to a distorted account of the functional similarities (or dissimilarities) between the two candidate regions. For example, Broca’s area’s domain-specific cognitive uses greatly differ from those of its hypothesized homologue in the brain of macaque monkeys (Petrides, Cadoret, & Mackey, 2005), as evidenced by the involvement of Broca’s area, but not of its monkey homologue, in the processing of linguistic and musical information. Focusing on this fact, however, can lead to a failure to appreciate the possibility that the same cognitive working zone might adequately capture the basic contribution of both areas to the cognitive processes in which they participate. In other words, whether or not two brain regions in two different species will appear to be functionally homologous may depend on which mode of functional specification is adopted.

Fig. 1 (Chapter 4, section I) describes four distinct kinds of form-‘function’ relationships which correspond to four different ways in which brain regions can be cognitively characterized. The foregoing analysis suggests that out of these four ways, the specification of cognitive working zones characterizes form-‘function’ relationships that are by far the most stable in the course of brain
evolution. This, in turn, suggests that the search for homologues of human neurocognitive structures in the brain of other animals for the purpose of building animal models of human cognition is more likely to succeed if it proceeds via the specification of cognitive working zones.

Going back to the vertebrate vision example, we can now see why the working zone approach is a good candidate for the introduction of homology thinking in the cognitive realm. If we can think of avian and mammalian vision as homologous despite their many different visual functions, it is because we can identify a system of cognitive working zones that they both share with the visual systems of common vertebrate ancestors.

Both of these conclusions are supported by recent data that show that the massive reuse and combination of existing neural structures may have played a significant role in cognitive evolution (Anderson, 2007b). In particular, they are supported by the following three findings: 1- most brain regions have many different cognitive uses in diverse informational domains, 2- evolutionarily older brain regions have more cognitive uses, and 3- more recent cognitive capacities engage more, and more widely scattered, brain regions. (1) is a direct consequence of the observation made earlier that in the course of evolution, significant changes to a region’s set of domain-specific cognitive uses will often occur in the absence of significant changes to its basic cognitive working. (2) is what would be expected if new and/or different versions of cognitive capacities (e.g. different kinds of vertebrate vision) are typically built, at least in part, upon a common set of core (and older) functional components. The older these core
components are, the more frequently they are likely to have contributed to the emergence of new capacities. And (3) is a consequence of the fact that the more recent a cognitive capacity is, the more likely it is that there were a large number of cognitive working zones ready to contribute to this new capacity at the time that it was evolving.

Perhaps we can end this section by noting that these results not only point to the positive contribution that the working zone approach can have on the development of animal models of human cognition, and on homology thinking in general in the cognitive realm, but also that they suggest that the specification of cognitive working zones, as opposed to other modes of functional specification, may be a better way to account for the continuity between animal and human cognition.

**IV. Conclusion**

I have been concerned, in this chapter, with the problem of mapping cognitive functions onto the brain. In particular, I have been concerned with the challenge posed by the presence of several different approaches to the functional specification of brain regions. I have tackled this challenge by first proposing a taxonomy of four different modes of functional specifications, which were then tested for their integrative value in three different contexts: the integration of models and theories within and across disciplines, the use of functional
neuroimaging data to distinguish between competing cognitive theories, and the building of animal models of human cognition. Out of these four modes, the specification of cognitive working zones turned out to be particularly valuable in all three contexts. Therefore, this approach to the functional specification of brain regions should be recognized as a productive way to achieve integration in cognitive neuroscience for the purpose of mapping cognitive functions onto the brain.
I. A further ambiguity in the meaning of ‘function’

The distinction between working and use, which is at the base of the proposed taxonomy of functional specifications, must be contrasted with another ambiguity in the meaning of the term ‘function’. This ambiguity is captured by the presence, in the philosophical literature, of two distinct approaches to defining functions. According to the first approach, the function of a trait is defined in terms of its evolutionary history (see e.g. Millikan, 1984; 1989; Matthen, 1988; Neander, 1991; Griffiths, 1993; Godfrey-Smith, 1994; Allen & Bekoff, 1995; Buller, 1998).

The core of this approach goes back to Larry Wright’s (1973) seminal analysis. Wright provides the following analysis of the meaning of ‘Z is the function of X’:

(a) X is there because it does Z, and

(b) Z is a consequence (or result) of X’s being there (p. 161).

In itself, this formulation does not introduce a historical element into the analysis of function ascription. However, in the case of biological entities, a historical element can be introduced via an appeal to natural selection. In this case, one can reformulate (a) as ‘X is there because Xs have done Z in the past’. One can
then explain the presence of a trait by an appeal to its past selection history.

Buller (1998) provides a typical formulation of the notion of biological function which falls within this etiological approach:

A current token of a trait $T$ has the function of producing an effect of type $E$ just in case, at some point in evolutionary history, there was selection for $T$ (over alternative items) because of its having produced effects of type $E$ (p. 506).

In essence, the etiological function of a biological entity is the effect for which it was selected. For instance, the function of the heart is to circulate blood, because the heart was selected for the circulation of blood (and not, say, the making of noise). This kind of adaptationist reasoning about functions also permeates cognitive science, and especially evolutionary psychology (more on this below).

According to the second approach to defining functions, the function of a biological entity is not defined in terms of its evolutionary history. The most influential account which falls within this ahistorical approach to defining functions says that the function of a given biological entity (organ, brain region, brain process, etc.) is the contribution that it makes toward the capacities of some larger entity of which it is part (Cummins, 1975). In simplified form, Robert Cummins offers the following analysis of ‘$Z$ is the function of $X$ in [larger system] $\mathcal{Y}$’.
(a) $X$ is a part of $Y$, and

(b) $X$ is capable of $Z$ within $Y$, and

(c) $Z$ contributes to some capacity of $Y$.

In other words, *cummins functions* are capacities of entities which contribute to the capacities of the larger entities of which they are part. For instance, the performance of rhythmic contractions is a cummins function of the heart which contributes to the capacity of the cardiovascular system to circulate blood. Similarly, according to Petrides (1996, 2005) model of lateral PFC function (Chapter 4, section III), the online maintenance and abstract recoding of multiple pieces of information are cummins functions of the dorsolateral PFC which contribute to the capacity of some larger brain systems (of which the dorsolateral PFC is part) to perform working memory operations.

Having distinguished etiological from cummins functions, it should be noted that the notion of etiological function is silent with respect to the working/use distinction. Etiological functions can in fact be separated into etiological workings and etiological uses. For instance, the performance of rhythmic contractions would be an etiological *working* of the heart, while the circulation of blood would be one of the heart’s etiological *uses* (similarly with etiological *cognitive* functions, which can be separated into etiological cognitive workings and etiological cognitive uses.) This is because etiological functions are conceived simply as selected ‘effects’, and as such they can be specified either as activities that particular entities perform (etiological workings), or as activities in which
particular entities take part, or participate (etiological uses).\(^{18}\)

The cummins functions of an entity, by contrast, can only refer to the workings of that entity, since the uses of that entity refer not to its own capacities, but to the capacities of some larger entity in which it is part (remember that cummins functions are capacities of entities which contribute to the capacities of the larger entities of which they are part.) For instance, the performance of rhythmic contractions is a working and a cummins function of the heart, but the circulation of blood—a use of the heart—cannot be a working nor a cummins function of the heart because it is the capacity of a larger entity (the cardiovascular system) of which it is part.\(^{19}\)

Note, however, that not all descriptions of the working of a biological entity will count as cummins functions. In order to count as such, the entity’s working must be specified in terms of its contribution to some capacity of a larger entity of which it is a part. This means that in order to count as a cummins function, the working of an entity cannot be specified completely in isolation. For example, a neurophysiological description of Broca’s area’s working solely in terms of the spontaneous firing patterns of its various types of neurons—and thus independently of any cognitive task, or of any account of its interactions with neurons in other brain regions—would not count as a cummins function of this area. By contrast, the description of Broca’s area as a hypersequential processor (see Chapter 4, section II) is couched in terms of how it organizes (or transforms)

---

18 One can also think of the etiological workings of particular entities as their internal selected effects, and think of their etiological uses as their external selected effects.

19 Notice that the circulation of blood is a working and a cummins function of the cardiovascular system, in so far as this description of the cardiovascular system’s activity helps explain the capacity (or capacities) of some larger entity of which it is part.
incoming sensory or motor information in preparation for further processing by other brain regions, and it is therefore a cummins function of this area.

So, not all workings of a biological entity are cummins functions, but all cummins functions of a biological entity are workings of that entity. And since a cognitive working zone is the working of brain region that best captures its basic contribution to the cognitive operations in which it participates, the working of a cognitive working zone must always count as a cummins function. Fig. 2 below outlines the four distinct notions of function which derive from the foregoing discussion of the historical and functional aspects of functional specification.

Fig. 3. Etiological versus non-etiological workings and uses. Four distinct notions of function. See text for details.

As this figure shows, etiological 'functions' can either be etiological workings or
etiological uses. By contrast, all cummins functions are non-etiologial workings, meaning that they are workings of (biological) entities that contribute to some capacity of a larger entity of which they are part.

We are now in a position to make an important clarification to the proposed taxonomy of functional specifications (Chapter 4, section I, Fig. 1). The proposed taxonomy is entirely ahistorical, in the sense that it is characterized by a non-etiologial interpretation of the notions of working and use. Therefore, just like cummins functions, the workings and uses that are attributed to brain structures do not have to be specified in relation to some past episode of natural selection. This means that brain structures may have a number of workings and uses that the etiological theorist would not recognize as proper functions20. For instance, the standard account of the fusiform face area’s (FFA) etiological function is that this brain region was selected for face recognition (Boyer, & Barrett, 2005). However, as noted in Chapter 2, we know that this region is also recruited when people have to perform fine-grained, within-category distinctions between grossly similar visual representations of non-face objects (Gauthier, Tarr, Anderson, Skudlarski, & Gore, 1999; Tarr & Gauthier, 2000). If we assume that the FFA was selected for face recognition and not for within-category distinctions more broadly, it would follow that the FFA has at least one non-etiologial use that the etiological theorists would not recognize as a function of this structure, but only as a mere effect.

But why, one might ask, should the functional specification of cognitive

---

20 Incidentally, ‘proper functions’ is a term that some use for ‘etiologial functions’ (see e.g. Millikan, 1984; Griffiths, 1993).
components be done non-etiologically, as it is now claimed, and not etiologically, as it is currently done in evolutionary psychology and, to a lesser extent, in other parts of cognitive science? After all, the working zone approach is committed to viewing the mind/brain as an evolved system, and it even considers the identification of homologues of human neurocognitive structures in the brains of other species to be a productive strategy for the specification of human cognitive working zones. The answer to this question is next.

II. Etiological versus non-etiological functional specification

Cognitive scientists all agree that the human mind/brain is the product of a long evolutionary history, and that evolution by natural selection has played an important role in shaping the architecture of human cognition. What cognitive scientists don’t all agree on is 1- how important was the contribution of evolution by natural selection, and 2- how best to approach the study of cognition within this general evolutionary perspective. In this section, I am concerned only with the second question.

It is useful to distinguish between two principal ways in which the evolutionary perspective plays out in cognitive science. The first way consists of using a comparative approach to the study of cognition, e.g. comparative neuroscience, comparative (neuro)psychology. A good example of this is the search for
homologues of human neurocognitive structures in the brains of other species. As we have seen in Chapter 5, section III, this strategy may play an important role in the successful identification of cognitive working zones. Let us call this approach the _comparative approach_.

Note that this approach can be pursued entirely non-etio logically. That is, the comparative approach to (neuro)psychology and cognitive neuroscience is _not_ committed to etiological functional specification.

The second principal way in which the evolutionary perspective plays out in cognitive science is through the methodological assumption that several (and perhaps most) cognitive components of the human mind/brain were designed by natural selection to solve _specific_ adaptive problems that were faced by our ancestors (see e.g. Carruthers, 2006; Pinker, 1997; Sperber, 1994; Tooby, & Cosmides, 1992; 2005). Evolutionary psychology is generally committed to this methodology, but the same kind of adaptationist reasoning also permeates, to a lesser extent, other fields of cognitive science. Let us call this approach the _adaptationist approach_.

Note that the comparative and the adaptationist approaches, as presently defined, are not incompatible. Indeed, etiological functional specification can be done (and is often done) in conjunction with the methods of comparative (neuro)psychology and cognitive neuroscience (see e.g. the articles in Buss, 2005).

One characteristic of the adaptationist approach is its commitment to etiological functional specification. For the adaptationist theorist, the ‘function’ of
a particular cognitive component is the effect for which it was selected. A consequence of this commitment to etiological functions is a further commitment to the notion of *proper* informational domain of a cognitive component (Sperber, 1994). The notion of proper informational domain is distinct from the notion of *actual* informational domain in that the *proper* informational domain of a cognitive component is the type of stimuli, or inputs, it is its etiological function to process, while the actual informational domain of that component is the set of all stimuli, or inputs, on which it operates. For example, the adaptationist theorist will claim that the *proper* domain of the FFA is the set of all faces, even if this cortical region’s *actual* domain also appears to include other kinds of grossly similar objects the visual recognition of which requires the performance of fine-grained, within-category distinctions.

The forgoing discussion of the distinction between the comparative and adaptationist approaches allows us to make an important qualification to the proposed framework with respect to its evolutionary commitment. The working zone approach is entirely non-adaptationist, in the sense that it does not presuppose that cognitive working zones were designed by natural selection to solve *specific* adaptive problems that were faced by our ancestors. Instead, the working zone approach’s commitment to evolutionary thinking is reflected mainly in its endorsement of the comparative approach. This, of course, is perfectly compatible with the possibility that most, if not all, cognitive working zones were at least in part designed by natural selection. The working zone approach is non-adaptationist not because there is any doubt that cognitive working zones are
adapted structures, but because of the counterproductive nature of adaptationist thinking when it comes to specifying the basic components of human cognition. I spend the rest of this chapter arguing for this last claim.

To begin with, there is a problem with the logic of the reasoning underpinning the adaptationist approach to functional specification. Tooby and Cosmides (2005) give the following account of the method for discovering new ‘neurocomputational systems’ within the adaptationist framework:

(1) Start by describing an adaptive problem encountered by human ancestors, including what information would potentially have been present in past environments for solving that problem.

(2) Develop a task analysis of the kinds of computations necessary for solving that problem, concentrating on what would count as a well designed program given the adaptive function under consideration.

(3) Based on such task analysis, propose a neurocomputational system that could implement these computations and which might have evolved by natural selection.

(4) Test for the presence of that system using the most appropriate experimental methods, including cognitive, social, and developmental
psychology, cognitive neuroscience/neuropsychology, experimental
economics, and cross-cultural studies (adapted from Tooby & Cosmides 2005,
p. 28).

The problem with this method is how one goes from (2) to (3). Even if one assumes that (1) describes a real problem encountered by human ancestors and that the task analysis developed in (2) accurately describes the kinds of computations that humans actually perform, it simply does not follow that one should be looking for the existence of a dedicated neurocomputational system specifically evolved to perform these computations. The kind of analysis going into (2) can be valuable if it suggests plausible ways in which humans (or their ancestors) came to have a particular cognitive capacity. But cognitive capacities and neurocomputational systems are two different things, and the kind of evidence required to establish the one is quite different from the kind of evidence required to establish the other. To illustrate this point, consider the example which Tooby and Cosmides (2005) offer as a demonstration of how the adaptationist method can be used to discover new neurocomputational systems:

Avoiding the deleterious effects of inbreeding was an important adaptive problem faced by our hominid ancestors. The best way to avoid the costs of inbreeding is to avoid having sex with close genetic relatives. This, in turn, requires a system for distinguishing close genetic relatives from other individuals: a kin detection system which computes a kinship estimate for each
individual with whom one lives in close association (p. 29, emphasis added).

In this passage, Tooby and Cosmides assume that the emergence of a capacity for kin detection, requires the emergence of a dedicated neurocomputational system specifically evolve to perform this task. Their analysis, however, is limited to a decomposition of the analyzed capacity into a number of sub-capacities which together allow the alleged system to 'compute a kinship estimate for each individual with whom one lives in close association'. For example, they mention the capacity to represent the ‘cumulative duration of childhood coresidence’; the capacity to recognize what they call ‘maternal perinatal association’, i.e. the direct observation of one’s own mother caring for another infant; and the capacity to form an ‘olfactory signature indicating similarity of the major histocompatibility complex’ (p. 29). In fact, none of the experimental data that they offer in support of this particular task decomposition (against other proposals of how humans perform kin detection) are the kind of data that can provide an adequate basis for inferring the existence of dedicated neural resources specialized for kin detection. In particular, none of the data are double dissociation data, either of the cognitive psychological, neuropsychological, or functional neuroimaging kind\(^{21}\). Therefore, even if their task analysis turns out to be right, such that the sub-capacities composing human kin detection are exactly as they claim, we still would have no good reason to think that there should be a dedicated neurocomputational system specifically evolved to perform this cognitive

\(^{21}\) See Chapter 1, section I and II for a discussion of the dissociation logic in cognitive science (see also the discussion below).
Another way to put this point is to compare the transition from (2) to (3) to the functional modularity inference discussed in chapter 2. Both kinds of reasoning rest on some version of what we have called the functional modularity assumption, i.e. the idea that the architecture of human cognition largely consists of a configuration of (relatively) functionally independent mechanisms, and that each of these mechanisms operate over a distinct and relatively specific domain of cognitive information (Chapter 1, section II). In addition, both kinds of reasoning are used to hypothesize the existence of cognitive modules. The crucial difference, however, between the two is in the kind of evidence on which they primarily based. On the one hand, the functional modularity inference is based primarily on dissociation data of various kinds (behavioral, neuropsychological, neuroimaging). The use of this particular kind of evidence, despite its own limitations (Chapter 2), constitutes the main driving force behind this methodological tool for inferring the existence of functionally independent cognitive mechanisms in the human mind/brain. On the other hand, as can be seen from the example just mentioned, support for the transition between steps (2) and (3) of the adaptationist approach to the discovery of new neurocomputational systems does not require the use of dissociation data of any kind. In fact, what seems to be the main driving force behind this transition is nothing more than the adaptationist theorists’ commitment to a particularly strong (and highly controversial) version of the functional modularity assumption, namely, the massive modularity hypothesis. We therefore have a good reason to
be worried about the use of this approach for the purpose of discovering the basic components of human cognition.

Note that in some cases, adaptationist theorists do make use of dissociation data as part of the empirical testing for the existence of a proposed neurocomputational system. For instance, Boyer and Barrett (2005) argue that the adaptationist account of the face recognition module hypothesis provides the best explanation of this system’s specificity, and they mention several different kinds of dissociation data in support of this hypothesis. In such cases, however, it is the dissociation data, and not the adaptationist scenario, that are providing the bulk of the support for the functional independence claim—i.e. the postulation of a dedicated neurocomputational system specialized for face perception/recognition. Thus, while it is true that in such cases the adaptationist approach may contribute to the discovery and/or functional specification of cognitive modules (the primary motivations for adopting the approach), the main contribution of adaptationist reasoning pertains to its role in task analysis, not in providing support for the existence of particular basic components of human cognition, let alone in providing support for the existence of cognitive modules.

A second problem with using the adaptationist approach for the purpose of discovering the basic components of human cognition has to do with its commitment to the specification of proper informational domains. We have already seen in Chapter 5 that an emphasis on domain-specific functional specification (i.e. form-use or form-working relationships) as opposed to
specification according to operation type (particularly cognitive working zones) can prevent integration in cognitive science. Specification according to proper informational domains worsens this problem.

To see this, let us first review the three contexts in which domain-specific functional specification can prevent integration in cognitive neuroscience. It can prevent the integration of independent lines of research pertaining to different informational domains (Chapter 5, section I); it can prevent the effective use of functional neuroimaging for the development and evaluation of cognitive theories (Chapter 5, section II); and it can prevent the successful identification of homologues of human neurocognitive structures in the brains of other species, thus preventing the successful development of animal models of human cognition (Chapter 5, section III). Thus, to the extent that the adaptationist approach is committed to domain-specific functional specification, it is susceptible to the same criticisms that were directed at functional specification that operates primarily via the specification of form-use and form-working relationships.

An emphasis on the specification of proper informational domains amplifies these three integration problems. The main reason for this is that in many (if not most) cases, the proper domain of a brain region’s cognitive activity will be significantly more narrow than its actual domain (see below for some examples). And the more narrowly specified the cognitive domain of a brain region is, the harder it is to identify its basic contribution to the cognitive operations in which it participates.
In Chapter 5 section I, we saw how an emphasis on the specification of form-use and/or form-working relationships within several lines of research can yield a collection of disparate domain-specific cognitive components. The proposed solution was to focus instead on the specification of cognitive working zones. This was illustrated with two examples of how the specification of cognitive working zones can help integrate several apparently incompatible specifications of the ‘function’ of a particular brain region (the two examples were 1- the involvement of Broca’s area in several linguistic and non-linguistic cognitive operations, and 2- the involvement of parts of the intraparietal sulcus in both numerical and spatial cognition.) This particular strategy exploits evidence of the involvement of a given brain region in more than one (or in a wide) informational domain to help zoom in on its basic contribution to the cognitive operations in which it participates. The adaptationist theorist, by contrast, focuses on the specification of proper (i.e. narrower) informational domains, and this makes it harder for this kind of integration to take place.

In Chapter 5 section II, we saw that the reverse neuroimaging inference, when it is legitimate and strong, can be a very useful tool for the testing and development of cognitive theories. As we also saw, the degree to which a given brain region is selectively activated by a cognitive operation is the greatest determinant of the strength of this inference. This means that the more narrowly specified the cognitive domain of a brain region is—e.g. specifying FFA’s domain as faces, as opposed to the wider domain consisting of grossly similar objects the recognition of which necessitates the performance of fine-grained, within-
category distinctions—the weaker any reverse inference involving that region is likely to be. And since the proper domain of a brain region is, at least in most cases, likely to be significantly more narrow than its actual domain, it follows that a reverse inference involving that region is likely to be significantly weaker when it operates within the adaptationist framework.

In Chapter 5 section III, we saw that the functional specification of human brain regions often depends on the identification, in the brains of other species, of homologous neurocognitive structures. As we have seen, this task is complicated by the fact that in the course of evolution, the set of cognitive uses of a brain region, and similarly the informational domain over which it operates, is likely to undergo significant changes. This is a problem because it means that the domain of operation of a human brain region will often differ considerably from the domain of operation of its homologue in the brain of another species, making it harder to discover important functional similarities between the two structures. One solution to this problem, as it was suggested, is to adopt the working zone approach, i.e. to aim at specifying the basic cognitive working that two candidate homologous regions might share. Moreover, we saw that the specification of cognitive working zones is made easier by considering the wide (i.e. actual) domain of each region (recall that specifying Broca’s area as a hypersequential processor depended on the consideration of its common involvement in various linguistic and non-linguistic operations.) But the adaptationist theorist, for her part, tends to favor functional specification according to proper, and therefore narrower, informational domain, and this makes it harder for her to accomplish
A third, and last, problem with using the adaptationist approach for the purpose of discovering the basic components of human cognition is that our knowledge of the particular evolutionary history of a brain region (or cognitive capacity) is generally very limited. In particular, we almost never have a solid understanding of what the proper informational domain of a brain region is, and in the absence of that kind of knowledge, the adaptationist approach loses its force (at least a general strategy for the discovery of the basic cognitive components of human cognition.)

To see this, consider again the discussion related to the supposed module for face recognition (Chapter 1, section II). The adaptationist theorist might be tempted, based on 1- several pieces of evidence which point to the relative specificity of the fusiform face area (FFA) for face recognition tasks, and 2- the likely selective advantage conferred upon human individuals of recognizing other humans by their faces, to conclude that the FFA constitutes a ‘module specialized for face perception’ (Kanwisher, McDermott, & Chun, 1997). She might also argue that the findings that the FFA is also active when experts on other objects—cars, birds—are engaged in identifying these objects (Gauthier, Tarr, Anderson, Skudlarski, & Gore, 1999; Gauthier, Skudlarski, Gore, & Anderson, 2000) are compatible with the claim that FFA is the site of a subprocess specialized for face recognition, but recruited for other purposes as well. In other words, it could be that face recognition skills, innate though they are, are
transferable to other tasks.

The problem with this proposal is that it is not more plausible than several alternative proposals that are based on the same kind of reasoning, but which do not construe the FFA as a face recognition module. To see this, notice first that what the expert object recognition studies suggest is not that the FFA is recruited for nonface recognition tasks only in a few people, i.e. the so-called car and bird experts. In fact, the expertise effect (increased activation) in the FFA is observed with novel nonface objects in any subject after only a few hours of being trained at recognizing these novel objects at the individual level (Gauthier, Tarr, Anderson, Skudlarski, & Gore, 1999). This, in turn, indicates that one would expect the FFA to be recruited whenever people have to perform fine-grained, within-category distinctions between grossly similar visual representations of any kind of familiar objects.

Now, it is easy to think of several different kinds of nonface objects that it would have been selectively advantageous for our ancestors to recognize at the individual level. For example, several species of terrestrial mammals use individual trees and individual pieces of landscape (hills, bushy areas, etc.) as food caches, or as landmarks for navigation. This behavior, in turn, requires the capacity to perform fine-grained, within-category distinctions between grossly similar visual representations of these environmental items. One could therefore conclude, following the above adaptationist reasoning, that such species (e.g. squirrels) must be equipped with a food cache, and/or landmark recognition module, which might very well be a homologue of the FFA in humans. Moreover,
and this is the important point, any of these special recognition skills could have been transferred, throughout mammalian evolution, to other recognition tasks, including face-recognition tasks. Under this scenario, therefore, it might very well be that a homologue of the FFA first emerged for reasons that had nothing to do with face-recognition, and it is only much later that primates (or a more ancestral species) began to use it for that purpose.

The point of this example is to show that we do not, at this point in time, possess sufficient knowledge of the evolutionary history of the mammalian visual system to prefer any one of several plausible adaptationist scenarios concerning the proper domain (if any) of the FFA. The working zone approach, by contrast, avoids this problem altogether by attempting to capture the basic (i.e. domain-neutral) contribution of the FFA to any recognition task in which it is involved. In addition to not having the problem of choosing between plausible alternative adaptationist scenarios, this approach has at least one other important advantage over the adaptationist approach. This advantage is that concentrating our efforts on specifying the FFA’s basic cognitive working may actually facilitate the search for a potential homologue of this area in the brain of other species (as discussed in Chapter 5, section III), which can in turn help advance our understanding of the evolutionary history of the cognitive capacities in which it participates.
Philosophy of Science, 62, 609-622.


Bouillaud, J.-B. (1825). Recherches cliniques propres à démontrer que la perte
de la parole correspond à la lésion des lobules antérieurs du cerveau, et à confirmer l'opinion de M. Gall sur le siège de l'organe du langage articulé.

*Archives Générales de Médecine*, 8, 25-45.


Fiebach, C. J. & Schubotz, R. I. (2006). Dynamic anticipatory processing of


Henson, R. (2005). What can functional neuroimaging tell the experimental


prefrontal lesions in monkeys. *Brain Res*, 143(2), 313-323.


Page, M. (2006). What can't functional neuroimaging tell the cognitive


Plaut, D. C. (1995). Double dissociation without modularity: evidence from


