CHAPTER 20

THEORETICAL ISSUES IN COGNITIVE PSYCHOLOGY

Tony Stone and Martin Davies

1 INTRODUCTION

In the preceding chapters you have met a wide variety of theories of our cognitive capacities. In addition to debates over competing theories of specific capacities (e.g. of visual perception, autobiographical memory, or reasoning), cognitive psychology has seen debates that range more widely across the different areas of cognition. These concern the key concepts and explanatory strategies that are used in modelling cognition. This chapter introduces you to some of these debates. Before reading on, try Activity 20.1.

In doing Activity 20.1 you may have noticed that nearly all the cognitive theories discussed in this book make use of the notion of *mental representation* (in addition to the notion of *mental processing* mentioned in Activity 20.1). The Bruce and Young (1986) model of face recognition discussed in Chapter 4, for example, contains a component labelled *face recognition units* (FRUs): these are stored mental representations of the faces with which you are familiar. Similarly, models of spoken word recognition (Chapter 6) make use of the notion of a *mental lexicon* – this can be thought of as a store that contains mental representations of all the words that you know and representations of various properties of those words (e.g. their meanings). In the study of memory, there are thought to be mental representations of the episodes in one's life (in episodic memory) and of the 'know how' that one has (in procedural memory). In each of these examples, the idea is

✱ ACTIVITY 20.1

Look through the notes relating to theories or models of cognition that you have made on previous chapters and try to identify themes and concepts that recur.

COMMENT

You might start by simply making a list of all the theories or models that have been discussed in the previous chapters and then look for similarities and differences between them. If you look, for example, at the dual-route cascaded (DRC) model of word recognition (Chapter 6), you might note that there are two routes for the pronunciation of written words. One route involves assembling the phonological representation of the word using spelling-sound rules (also called the 'rule-based' route in Chapter 6), and the other route links the written form of the word with a stored representation of its correct pronunciation (also called the 'lexical' route in Chapter 6). You might then investigate the extent to which other cognitive models involve these two distinct kinds of *mental processing*.

that there is something internal to the mind that encodes information about the external world and about one's experiences, past, present, or future.

A fundamental principle of contemporary cognitive psychology is that cognition involves the storage and processing of information and that this information processing is achieved by *the processing or transformation of structured mental representations*. There are two basic kinds of processing in evidence in contemporary cognitive models. Sometimes representations are transformed by *rules*. For example, in Marr's model of the early stages of visual processing (see Chapter 3) we find the successive transformation of mental representations by a kind of rule called an algorithm. Thus, the grey level description is transformed into the raw primal sketch by an algorithm that compares images that have been blurred to different degrees. Similarly, in the DRC model of word reading, the pronunciation of non-words such as SLINT 'requires the postulation of a system of rules . . . loosely referred to as the letter-sound rule system' (Coltheart *et al.*, 1993, p.590). On other occasions information processing is conceptualized as the *transmission of activation* from one representation or group of representations to others. Whilst this kind of information processing is particularly characteristic of connectionist models of cognition (see Chapter 19), it is also found in more traditional models. In the Bruce and Young model of face recognition, there are connections between the FRUs and the person identity nodes (PINs) with activation flowing from an active FRU to the associated PIN.

The idea that cognition can be understood in terms of the rule-guided transformation of mental representations is at the heart of the computational model of the mind that has been dominant throughout most of the history of cognitive psychology. The first theoretical debate that we shall examine, in Section 2, concerns whether human information processing involves rules and representations to the extent proposed by many of the cognitive models you have studied.

A further thing you may have noticed when doing Activity 20.1, though you might have thought it too obvious to mention, is that there are different models for different cognitive capacities. Thus, there are separate models for spoken word recognition and visual word recognition (Chapter 6), for object recognition and face recognition (Chapter 4), for episodic memory and autobiographical memory (Chapters 8 and 17), and for speech comprehension and production (Chapters 6 and 7). This is not just a matter of practical convenience reflecting the specialized interests of different cognitive psychologists. Rather, it reflects the idea that the mind itself is composed of relatively independent, special-purpose systems that carry out specific information-processing tasks – that is, the idea that the mind is modular. Section 3 of this chapter is devoted to debates about modularity.

Debates on these two themes – rules and representations and the modularity of the mind – have been central to recent theoretical work, but there are other important questions that have also been subject to intense discussion. For instance, it is one of the attractions claimed for many connectionist models that they are more *brain-like* than traditional models. Yet it is noticeable that, in most of the preceding chapters, less is said about the way in which the various models of cognitive functioning are actually implemented in the brain. It may seem that there is only a poorly understood connection between, on the one hand, cognitive modelling and theories about cognitive processes and, on the other hand, the actual study of the physical brain. Nonetheless, much of the most compelling data for cognitive models over the past 35 years or so has come from the study of people who suffer from various kinds of cognitive impairment resulting from brain injury (Ellis and Young, 1998, give a textbook introduction to this work). To take just one example from many, the study of people with prosopagnosia (a specific inability to recognize once-familiar faces) has provided important evidence for the development of the Bruce and Young model of face recognition. This interaction between the study of normal and disordered cognition raises searching theoretical questions. For example, how exactly are inferences about normal cognitive processes made from cases of impaired cognitive functioning? These questions are taken up in Section 4.

The importance of neuropsychological evidence (Chapter 13) and the rapid development of neuroimaging techniques (Chapter 14) both highlight the need for a clear account to be given of the relationship between the kinds of models discussed in this book and the study of the brain. Will cognitive psychology ultimately reduce to – or will it perhaps be replaced by – neuroscience? These issues will be the topic of Section 5.

SUMMARY OF SECTION 1

- A fundamental principle of contemporary cognitive psychology is that cognition involves the processing of structured mental representations.
- Two kinds of processing are in evidence in cognitive models: rule-guided processing and the transmission of activation.
- Cognitive modelling reflects the idea that the mind is composed of modular systems.
- Inferences are made from neuropsychological case studies to models of normal cognitive functioning.
- There is a need to clarify the relationship between cognitive models and neurobiological models of the brain.

2 COMPUTATION AND COGNITION

One question that might occur to you as you near the end of this book is whether there is a *general* theory of the mind (or of mental functioning) that can be distilled from the preceding chapters. After all, one of the aims of science is to try to find general theories for seemingly disparate phenomena. One proposal for such a general theory has been influential throughout the history of cognitive psychology – indeed, it was present at its inception. This is the view that the mind is a computational device; that cognition is computation.

2.1 The computational model of the mind

We have already met the basic notions that lie behind this approach in discussing Activity 20.1 – it is the approach that sees cognition as the rule-guided transformation of structured mental representations. We refer to this idea as the computational model of the mind (CMM).

An excellent example of this idea in action is the DRC model of word reading (Coltheart *et al.*, 1993, 2001) that you met in Chapter 6, and that was mentioned in Section 1. This model proposes that there is an assembled phonology route where the pronunciation of regular words and non-words is computed via spelling-sound rules. The pronunciation of the non-word SLINT, for instance, is arrived at through the application of the rules that S is pronounced /s/, L is pronounced /l/, I is pronounced /i/, and so on. It is important to be clear that these spelling-sound rules are not thought merely to be summary *descriptions* of the ways in which English regular words and non-words happen to be pronounced, but are meant to be part of the *causal* story of how people produce pronunciations of regular words and non-words of English. (Strictly speaking, familiar regular words will generally be pronounced via the lexical route, but they can also be pronounced via these spelling-sound rules.)

Let us compare the causal role played by spelling-sound rules in generating the pronunciations of words and non-words (according to the DRC model) with the role played by a rule that merely describes the operation of a system. Consider, for example, Ohm's law (this example is taken from Gallistel, 2001). Ohm's law states that in any electrical circuit, $I = \frac{V}{R}$ (where 'I' represents current, 'V' represents voltage, and 'R' represents resistance). The symbols 'I', 'V', and 'R' refer to measurable properties of an electrical circuit, and we can manipulate the symbols and make predictions that we can then test by measurement. For example, we can deduce from $I = \frac{V}{R}$ that $IR = V$, and then we can measure current and resistance, compute the numerical product $(I \times R)$, and see if the result matches what we find when we measure voltage (V). But the electrical circuit itself does not need to do the same computation. Ohm's law accurately *describes* the behaviour of electrical circuits, but computation guided by this law is not part of the *causal* story of why an electrical circuit behaves in the way it does. The electrical circuit itself does not contain *representations* of the current, voltage, and resistance, or a representation of the law that describes their relationship.

It is essential to the CMM that cognition is computation in the sense that it involves the *rule-guided processing* of structured mental representations. For processing to be *guided* by a rule, the rule must be part of the causal story of how the processing takes place.

Why is this approach called the *computational* model of the mind? The reason is that the digital computer is an example of a physical device that can process information by transforming symbols via a program – a set of rules stored in memory. Thus, the thought lying at the root of the computational view is that the brain (a physical device) processes information in the way that a computer does, in so far as human information processing involves transformation of mental representations and this transformation is guided by rules (akin to the computer program). It is important to recognize that this is a *model* of the mind. Models aim to capture what is fundamental to the thing being modelled – just as a geographical map (another kind of representation) attempts to capture what is fundamental to the terrain being mapped. So the CMM only aims to capture those aspects of the mind that are thought to be fundamental to information processing. Thus, the fact that a computer is made of silicon, wires, metal, and plastic (the computer hardware), whereas human information processors are made of flesh and blood (our hardware or, perhaps better, wetware), is not an objection to the CMM, because those aspects of a computer are not part of the model (and reasonably so given that the aim is to model the human mind, not human tissue).

But, you might wonder, shouldn't a model of the mind at least try to model that part of the human body – the brain – wherein mental processing occurs? Shouldn't we model the physical networks of neurons and the transmission of physical electro-chemical signals that actually implement information processing? It is a key aspect of the CMM that it does not try to model those neural processes as such. It does aim to provide a model of the brain, but at a level that is more abstract than the physical level. The CMM models 'the mind as the software of the brain' (Block, 1995). It is a matter of debate, of course, whether this degree of abstraction, this failure to model the brain's physical processes as such, is really a virtue, rather than a vice, of the CMM.[1]

2.2 Connectionist modelling

We turn now to the first of the theoretical debates that we want to consider in this chapter – connectionist objections to the CMM. Connectionism (also known as 'parallel distributed processing' ('PDP') and 'neural network modelling') is a style of psychological theorizing and model building that re-emerged in the middle of the 1980s, and has become massively influential in cognitive psychology since then.

Though there are many types of connectionist model, typically a model will be a network composed of three layers of artificial neurons or *units* – an input layer and an output layer with a layer of hidden units sandwiched between (Chapter 19). There are multiple *connections* between adjacent layers, and these connections transmit *activation* from input units to hidden units, and from the hidden units to output units. The level of activation passed from units in one layer to units in the next layer is a function of the activation of the units in the first layer (the units that are passing on activation) and the strength of the connections between these units and the units in the next layer (the units that are receiving activation). Through a process of adjusting the strengths of connections – the *weights* on the connections – such models are trained to associate patterns of activation across units in the input layer with patterns of activation across units in the output layer.

Patterns of activation in the input and output layers are the network's input and output *representations*. For example, written words of English might be represented by patterns of activation in the input layer and the output patterns of activation that the model is trained to produce might represent the pronunciations of those words. Networks of this kind have been developed as competitors to the DRC model of reading English words aloud (Seidenberg and McClelland, 1989; Plaut *et al.*, 1996). In the development of a connectionist model of reading aloud, the modeller has to decide how written words (made up of letters) will be represented by patterns of activation across the input layer. Will the representation of a word be made up of representations of letters? If so, will individual letters be represented by activation of single units, or by patterns of activation across several units? How will the order of the letters in a word be represented, so that BAT and TAB have distinct representations? Similar decisions have to be made about how pronunciations (made up of phonemes) will be represented by patterns of activation across the output layer.

A key aspect of connectionist models is that they can learn to associate patterns of input activation with patterns of output activation; they are not programmed in advance with an algorithm or rule that specifies what pattern should be associated with what. (There are various learning procedures, of which the most important is back-propagation of error. We need not go into the details here but see Chapter 19.)

At their most radical, connectionist modellers say that they aim to model human cognition in ways that dispense with rule-guided mental processing. For instance, McClelland and Patterson say that, in the connectionist approach to the psychology of language, 'cognitive processes are seen as graded, probabilistic, interactive, context-sensitive and domain-general. . . . Characterizations of performance as "rule-governed"

[1] It might be thought that the CMM must be wrong because we are conscious and computers are not. But this would be too quick. The CMM is a model and does not have to capture all aspects of human psychology. Perhaps a different model could deal with consciousness. However, it would be an objection to the CMM if failing to model consciousness meant that it thereby failed to model cognition – a claim made by John Searle (e.g. 1992).

are viewed as approximate descriptions of patterns of language use: no actual rules operate in the processing of language' (McClelland and Patterson, 2002b, p.465).

Connectionism thus attacks a fundamental aspect of the CMM. It claims that mental processing is typically not rule-guided.

2.3 The past-tense debate: connectionism versus the CMM

We are going to present the connectionist challenge to rule-guided mental processing by considering the so-called past-tense debate. First, we set the scene for the challenge by describing an approach that does conform to the CMM.

2.3.1 The words and rules model

The past-tense debate concerns how we should best understand the ability, possessed by all competent speakers of English, to form the past tense of English verbs. (The debate has also been concerned with how children *develop* the ability to form the past tense, but we leave such developmental questions to one side in this chapter.) This example is not as complex as some of the models and tasks you have met in this book, but it has been the subject of considerable debate, and the apparent simplicity of the phenomena being modelled allows one of the fundamental divides between the CMM and connectionism to be clearly seen.

Consider, then, the problem of how a competent speaker of English puts a verb into the past tense. The past tense of the overwhelming majority of English verbs takes the form VERB STEM + PAST TENSE MORPHEME. Thus the past tense of the verb TO HUNT is HUNT + ED (HUNTED), the past tense of the verb TO STROLL is STROLL + ED (STROLLED), the past tense of the verb TO JUMP is JUMP + ED (JUMPED), and so on. (Note that, although the written form of the past tense morpheme is the same in these examples, its phonology does vary.) However, there is a minority of verbs (around 180 of the most common in the English language) that do not follow this + ED pattern. These verbs form their past tense in seemingly irregular fashions. Thus the past tense of the verb TO GO is WENT (not GO + ED) and the past tense of the verb TO BUY is BOUGHT (not BUY + ED).

One way to model how people form the past tense has been developed by Steven Pinker and his colleagues (e.g. Pinker and Ullman, 2002a and b; Ullman *et al.*, 1997). This model – the **words and rules model** – 'claims that the regular-irregular distinction is an epiphenomenon of the design of the human language faculty' (Pinker and Ullman, 2002a, p.456). In outline, the model posits that two separate structures within the language faculty are responsible for the formation of the past tense: the lexicon and the grammar. When a past tense is to be formed, both the lexicon and the grammar are accessed in parallel. Verbs that have an irregular past tense will access the appropriate form in the lexicon. Verbs that have the regular past tense will access no past tense form in the lexicon, and the grammar component of the language faculty will therefore add the regular + ED ending or inflection. (This is, then, a *dual-route* model analogous to the DRC model of reading aloud, with which you are familiar.) Given that the lexicon and the grammar are accessed in parallel, the model posits an inhibitory signal from the lexicon to the grammar. Whenever an irregular past tense (e.g. BOUGHT) is accessed in the lexicon, this inhibitory signal then blocks the parallel rule-guided formation of an incorrect regular form (e.g. BUY + ED).

This model (see Figure 20.1) has been coupled with a hypothesis that draws on the distinction between declarative and procedural memory (Chapters 8 and 19). It has been proposed (Ullman, 2001; Pinker and Ullman, 2002a) that the lexicon is part of declarative memory and is subserved by temporal and temporo-parietal regions of the neocortex, and that the grammar component is part of procedural memory and is subserved by the basal ganglia and the areas of frontal cortex to which they project. This proposal links the words and rules model with neural structures that are known, from independent evidence, to be involved in language processing. Moreover, independently of the debate about the past tense, declarative memory is thought to be responsible for the retention of facts,

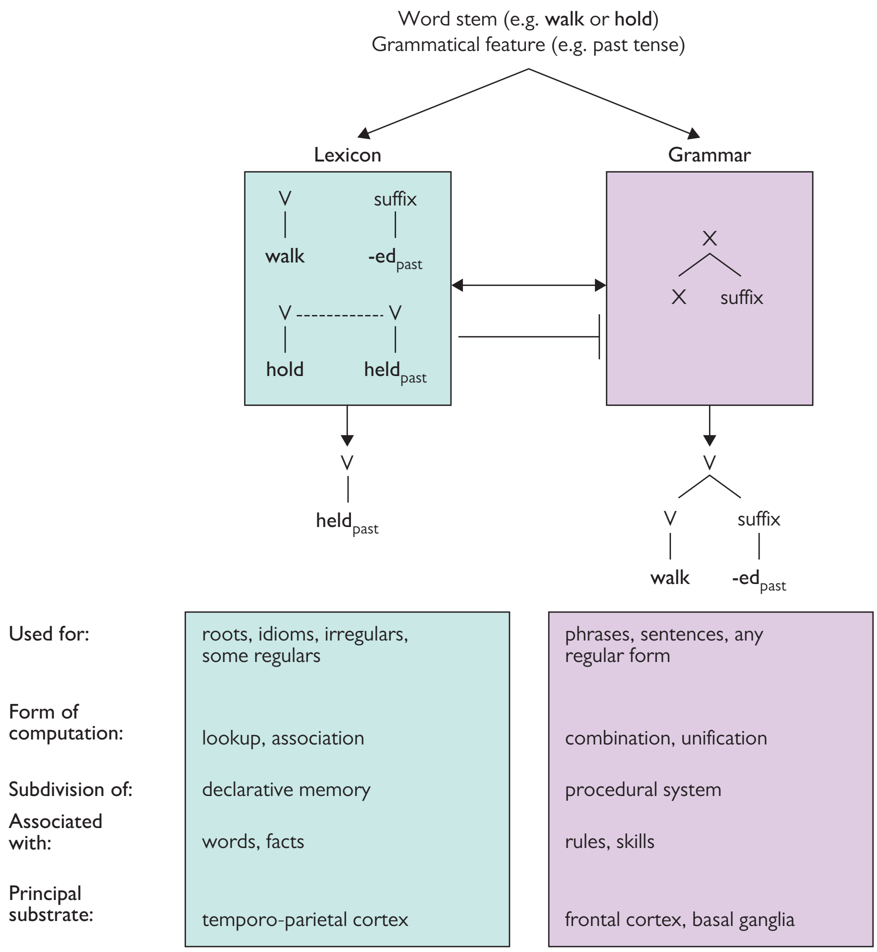

FIGURE 20.1 A simplified illustration of the words and rules theory and the declarative/procedural hypothesis. When a word must be inflected, the lexicon and grammar are accessed in parallel. If an inflected form for a verb (V) exists in memory, as with irregulars (e.g. *held*), it will be retrieved; a signal indicating a match blocks the operation of the grammatical suffixation process via an inhibitory link from lexicon to grammar, preventing the generation of *holded*. If no inflected form is matched, the grammatical processor concatenates the appropriate suffix with the stem, generating a regular form. *Source: Pinker and Ullman, 2002a, Figure 1, p.457*

and procedural memory responsible for the learning and control of motor skills, including skills that require sequencing. The thought, then, is that retention of an irregular past tense is the retention of a fact, whereas the construction of regular past tense forms depends on retention of a procedure.

2.3.2 Connectionist modelling of the past tense

Before reading on, try Activity 20.2.

Connectionist models aim to account for cognitive capacities without relying on rule-guided mental processing. How can this be achieved for the past tense? The first connectionist model of past-tense formation was developed by Rumelhart and McClelland (1986). This was a **pattern associator**, a simple network with just two layers of units (input and output units, but no hidden units) and one layer of connections. Patterns of activation across input units represent verb stems; patterns of activation across output units represent past-tense forms. (These are representations of the phonology, rather than the orthography (written form), of verbs and their past-tense forms.) The network was trained to associate the correct

✱ ACTIVITY 20.2

Try to identify the aspects of connectionist modelling that make it different from many of the other kinds of psychological models you have met in this book. For each difference try to find a specific connectionist model that illustrates the difference. In addition to this book, a good – freely available – web-based resource that will help you is the entry on connectionism (Garson, 2010) in the *Stanford Encyclopedia of Philosophy* (http://plato.stanford.edu/entries/connectionism/).

COMMENT

One difference you might find mentioned is that connectionist models are more 'brain-like' than models that are characteristic of the CMM. In what ways are connectionist models more 'brain-like' and are they very 'brain-like'?

past-tense form to both regular and irregular English verbs. Yet it did this with just a single route. All verbs, both regular and irregular, were processed via the same set of weighted connections from input units to output units. This appears to show that connectionist models are capable, in principle, of accounting for what looks like rule-guided behaviour in a way that does not require the use of a rule!

Rumelhart and McClelland's model learned to associate the appropriate past-tense form with each verb stem by relying on the statistical regularities contained in the training data. No rule was programmed into the model, and the trained model did not learn the task by representing and storing its own rule, such as 'add -ED to regulars'. How then does the model perform the task? Understanding this – and the same goes for any connectionist model – requires exploring in detail the activations of units and the weights on connections in the trained network. It should not be assumed, for instance, that the Rumelhart and McClelland model forms the past tense of all verbs in the way that the words and rules model forms the past tense of irregular verbs. It should not be assumed, that is, that the correct past-tense form of every verb is stored in a kind of declarative memory and then addressed by the appropriate verb stem.

The Rumelhart and McClelland (1986) model came in for some severe criticism from defenders of traditional rule-based theories (e.g. Pinker and Prince, 1988) but subsequent connectionist models have been able to meet some of those criticisms (e.g. Plunkett and Marchman, 1993; but see Pinker, 2006). What Rumelhart and McClelland (and other connectionist modellers) have shown, at least, is that there is an alternative to dual-route, 'words and rules' style models. But does the single-route connectionist alternative provide a better account of the available data than the dual-route words and rules model?

McClelland and Patterson (2002a) argue that connectionist models do provide a better fit with the data. They argue that the English past tense is a *quasi-regular* domain: in all but a few cases, past-tense forms that do not conform to the regular rule still display aspects of the regular pattern. Thus, McClelland and Patterson point out, the vast majority of English verbs with a so-called irregular past-tense form fall into one of nine groups, each of which exhibits aspects of the regular past tense. For instance, there is a group of verbs, including SAY, DO, TELL, and FLEE, whose past tenses are SAID, DID, TOLD, and FLED. Verbs in this cluster form the past tense by adding the /d/ sound as in the regular past tense but with a vowel adjustment to the stem. Another group, including BRING, CATCH, SEEK, TEACH, and THINK, form their past tenses by replacing the final consonant cluster with /t/ (the sound of the regular past tense following an unvoiced consonant, as in JUMPED) and adjusting the middle vowel to /O/ (sounds like 'aw'); thus BROUGHT, CAUGHT, and so on.

McClelland and Patterson claim that this quasi-regularity is not well explained by the words and rules model, in which formation of irregular and regular past-tense forms depends on two quite distinct mechanisms – the lexicon (words) and the grammar (rules). Quasi-regularity is, however, captured by connectionist models that have a single system for the formation of the past tense. McClelland and Patterson explain that their network can make the transformation KEEP → KEPT by simply:

> *adjust[ing] the activations of the output units representing the vowel, something the network will have learned to do on the basis of experience with* keep *and its neighbours* creep, leap, sleep, sweep, *and* weep. *The network uses the same connection-based knowledge that allows it to perform the regular mapping, and also taps into specific connections activated by the particular properties of keep to produce the vowel adjustment.*
>
> (McClelland and Patterson, 2002a, p.464)

In response, Pinker (2006) proposes a modified version of the words and rules model in which irregular past-tense forms are indeed stored in the lexicon, but the lexicon 'is not just a list of unrelated slots, but is partly associative'. Consequently, 'irregular verbs are predicted to show the kinds of associative effects that are well-modeled by pattern associators' (Pinker, 2006, p.223).

Pinker and Ullman (2002a; see also Pinker, 2006) draw attention to aspects of past-tense formation that they believe will be hard for connectionist models to deal with since their explanation lies in relatively deep linguistic principles. An example they provide is the way in which some usually irregular past-tense forms (e.g. STAND → STOOD) become regular in certain contexts such as GRANDSTANDED. These forms (regularizations of irregulars) occur due to linguistic principles concerning the formation of complex words of English. The verb TO GRANDSTAND is derived from a noun (A GRANDSTAND). Because of this, it is not possible for the irregular form (STOOD) stored in memory to be accessed – since that form must be accessed via a verb stem. Hence, no inhibitory block is sent from the lexicon to the grammar and the default rule is applied to form the regular past tense (GRANDSTAND + ED).

We shall return to the past-tense debate in Section 4.2.

2.4 Rules in connectionist networks

The leading idea of the CMM (the computational model of the mind) is that cognition involves the *rule-guided* processing of *structured* mental representations. If we are to understand connectionist modelling as a genuine alternative to the CMM, then we need to see how the information processing in connectionist networks departs from this idea. There is no doubt that *processing* takes place in networks and we have said that patterns of activation over input units, for example, are *representations* of items (such as words) in the network's task domain. In this section, we shall explore in what sense this processing might not be 'rule-guided' and, in the next (Section 2.5), in what sense the representations might not be 'structured'.

Rumelhart and McClelland draw attention to a difference between connectionist processing and rule-guided processing by saying that processing in a connectionist network does not require 'the postulation of explicit but inaccessible rules' (1986, p.218). According to the CMM, the rules that guide cognitive processes are often *inaccessible* to the person in whom those processes are taking place. We cannot simply introspect how cognitive processes work. (That is why we need to study cognitive psychology!) The CMM also allows that these rules, which guide cognitive processes, may be *explicit* in the sense that the rules are represented (in some format or syntax), are stored in memory, and have to be accessed before they can be used. A rule that is represented and stored, but is inaccessible to introspection, is sometimes said to be *tacitly known*. According to the CMM, there is a very clear distinction between having tacit knowledge of a rule and actually using that knowledge in processing. It is as clear as the distinction between having a cake recipe written in a book that is stored on the kitchen shelf and actually using the recipe to guide the processes of sieving flour, beating eggs, mixing ingredients, and so on, in order to make a cake.

In connectionist networks, rules are not explicit in this sense and there is a much closer relationship between having knowledge and using it. The processing in a network leads from patterns of activation across units in the input layer to patterns of activation across units in the output layer. The input patterns may represent written words and the output patterns may represent pronunciations; or the input patterns may represent verb stems and the output patterns may represent past-tense forms. If the network associates the correct pronunciations with written words or the correct past-tense forms with verb stems, this is because the weights on the connections have been adjusted (through a process of training such as back-propagation of error) to achieve this result. The trained network's knowledge of the task domain is thus embodied in the weights on the connections between units in adjacent layers. If the network's knowledge is in the weights then there is no need to access that knowledge before it can be used. When units in the input layer are activated, the activation is immediately passed forward from layer to layer, through the weighted connections, and a pattern of activation across the output units is produced. As McClelland *et al.* put it: 'Using knowledge in processing is no longer a matter of finding the relevant information in memory and bringing it to bear; it is part and parcel of the processing itself' (1986, p.32).

When we introduced the idea of rule-guided processing in the CMM, we said that rules are supposed to be part of the *causal* story of how a task is performed. For example, letter-sound rules are supposed to be part of the causal story of how the task of reading words aloud is performed. We contrasted this with the idea of rules as merely providing a summary *description* of patterns in the task domain. For example, it might be true of some domain of words that each word that begins with the letter B has a pronunciation that begins with the phoneme /b/. A successful model of the reading aloud task will perform in a way that *conforms* to this B → /b/ rule. But conformity to the rule does not tell us how the performance is actually achieved – it does not tell us whether the rule is merely *descriptive* or whether the rule is *causal*.

One possibility – in line with the CMM – is that the processing in the model is guided by explicit rules that are stored in memory, and have to be accessed before they can be used. An example of such a case is, again, the implementation of the assembled phonology route in the DRC model of reading. Coltheart *et al.* (1993) describe the model's use of spelling-sound rules as follows:

> *When confronted with a letter string for translation, it [the model] seeks to apply the rules to the string from left to right, starting with the longest possible rule that could accommodate that string. For the word* chip *it would start with four-letter rules, looking for a rule that maps the letters* chip *onto a single phoneme. No such rule will be found in the rule base. So a rule corresponding to the first three letters* chi *is sought; none will be found. The search for a rule for the first two letters* ch *[*ch → /tʃ/ *(sounds like* 'ch'*)] will, however, be successful.*
>
> (Coltheart et al. 1993, p.601)

You will not find explicit letter-sound rules in a connectionist model of reading aloud. But there is still a sense in which rules can figure in the causal story of how a connectionist model works and not just as a description of the network's performance. We can introduce this sense by considering a 'toy' connectionist network that takes representations of 25 written consonant–vowel pairs (built from five consonants and five vowels) and associates each of these with a representation of the corresponding pronunciation. For example, when given a representation of the written pair BA as input the network produces a representation of its pronunciation /bæ/ (sounds like 'ba' as in 'bat') as output. The network is represented in Figure 20.2.

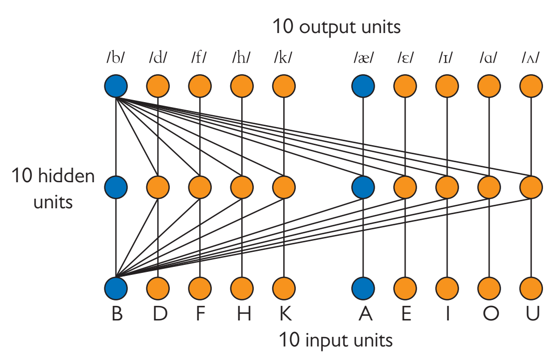

FIGURE 20.2 A connectionist model with implicit letter-sound rules (not all connections are shown).
Source: Based on Davies, 1995, Figure 3, p.181

This network associates input representations with output representations in a way that conforms to the B → /b/ letter-sound rule. How does the network achieve this performance? Examine the network shown in Figure 20.2 carefully. The input representation for a consonant–vowel pair such as BA involves activation of two units in the input layer (in this case, the consonant unit that represents B and the vowel unit that represents A). Whenever a pair beginning with B is presented to the system, the same input unit (representing B) is activated and, because the network conforms to the letter-sound rule, we know that the same output unit (representing the phoneme /b/) is activated.

Now, the important point to notice is that, whichever of the five pairs beginning with B is being processed, the causal explanation for the activation of the /b/ unit in the output layer is exactly the same. Activation is passed from the B unit to the leftmost hidden unit and on to the /b/ unit. So the five cases of conformity to the B → /b/ letter-sound rule have a *common causal explanation*. The explanation is grounded in a *causal common factor* – the passing of activation along the weighted connections from the B unit to the hidden unit and on to the /b/ unit. We could say that the B → /b/ rule is *implicit* in the network. The same, of course, goes for the nine other letter-sound rules that capture the pronunciations of these consonant–vowel pairs. This notion of a rule being implicit in a processing system requires more than just conformity to the rule. It requires that all the transitions that conform to the rule have a common causal explanation. But the notion of a rule being implicit in a system does not require that the rule is explicitly represented and stored and then accessed when it is needed.

Rules can figure in the causal story of how a connectionist model works because rules can be embodied implicitly in networks:

> *Is there, within the system [network], a component mechanism, or processor, or module that operates as* a causal common factor *to mediate all the input-output transitions that instantiate the pattern described by the rule? If so, then the rule is said to be implicit in the system (or the system is said to have implicit or tacit knowledge of the rule).*
>
> *(Davies, 1995, p.162)*

The point to take away from this discussion is *not* that all connectionist models embody implicit rules in this sense. There can certainly be networks whose performance conforms to rules, even though the input-output transitions that conform to a particular rule do not have a common causal explanation. When examining a connectionist model that performs a cognitive task, it is often useful to ask whether the performance is achieved by virtue of a structure of causal common factors that are the implicit analogues of explicit rules. If the network does not embody implicit rules, then we need to try to discover how the model does complete the task set for it.

2.5 Connectionism, structure, and compositionality

Connectionist modelling also challenges the style of *mental representation* that is characteristic of the CMM. This is a complex and (in parts) technical debate and we will briefly discuss just one issue.

The CMM holds that the mental representations processed by rules have **compositional structure** and that this structure explains a fundamental property of our cognitive systems – their **systematicity**. This latter property can most easily be introduced by reference to our linguistic abilities. Suppose I understand the sentences 'The dog bit the cat' and 'The cow jumped over the moon'. My understanding of these sentences is systematic in that I need to learn nothing more in order to understand the additional sentences 'The cow bit the dog' and 'The cat jumped over the moon'. This systematic understanding is possible because sentences are composed of words, and the same words, with the same meanings, can be deployed in different sentences to express new thoughts. It is this compositional structure of sentences that makes systematic understanding of a language possible.

Contrast this systematic understanding of our native language with what might happen when someone learns a second language from a phrase book. It is perfectly possible for a monolingual English speaker to learn to say the Romanian sentences 'O sticle cu vin roşu' and 'Un caiine cu dinţ mici' in appropriate circumstances and yet be unable to combine the words in these sentences in new ways to express appropriate new thoughts. Equally, an English speaker might learn to recognize and respond to utterances of those two sentences without being able to understand any other sentences of Romanian. You could achieve systematic understanding – going beyond those two sentences to use and understand additional sentences constructed from the same words – only if you knew what each word in the sentences meant and had some rudimentary grasp of the grammar of the language (how the words are put together). Sentences of Romanian do, of course, have compositional structure, just as sentences of English do. But knowledge of the meanings of a few phrases or sentences does not guarantee that you will become attuned to this structure (indeed, typically you will not). The CMM supposes, at least for the most part, that mental representations have compositional structure and that this structure is part of the causal story of how cognitive processes take place.

Connectionist modelling challenges this approach to mental representation. It is claimed, for example, that connectionist models can explain the systematicity of human cognition without appeal to mental representations with compositional structure (at least, in the sense of compositional structure used by the CMM).

Consider a simple example – this time in the domain of thought, rather than language. It seems to be a fact about human thought that if I am able to think the thought *New York is dangerous* and to think the thought *London is safe*, then I can also think the thought *New York is safe* and the thought *London is dangerous* (systematicity again). The capacity to think the first two thoughts seems to be intrinsically linked to the capacity to think the second two thoughts. The CMM explains the systematicity of thought by saying that the mental representations that underpin thinking have compositional structure, just as the sentences of a language have compositional structure. This is the *language of thought* hypothesis (Fodor, 1975, 1987; see also 2008).

It is certainly possible for connectionist representations to be compositionally structured. The network in Figure 20.2 represents the written consonant–vowel pair BA by activation of two units in the input layer, a consonant unit that represents B and a vowel unit that represents A. The consonant–vowel pair DE is represented by activation of a consonant unit that represents D and a vowel unit that represents E. Recombination of the resources in these two representations allows the network to represent the pairs BE and DA. Compositional structure allows systematicity of representation. In a similar way, a network might represent *New York is dangerous* by activation of a unit that represents the city of New York and another unit that represents danger. The thought *London is safe* might be represented by activation of a unit that represents the city of London and a unit that represents safety. Recombination of these representational resources would allow the network to represent *New York is safe* or to represent *London is dangerous*. Once again, compositional structure explains systematicity, just as it does according to the CMM.

Although connectionist modelling of cognitive processes allows for the possibility of representations with compositional structure, it does not require compositionality. There might be no pattern of activation across units that is present in every representation of a consonant–vowel pair beginning with B. There might be no pattern of activation that is shared by all of a network's representations of thoughts about New York. In a limiting case, each consonant–vowel pair, or each thought, might be represented by activation of a distinct unit or pool of units.

It is at this point that the challenge to the CMM becomes clear because, even without compositional structure, systematicity can, in a way, still be preserved. The modeller just needs to make sure that, if there is an input unit that represents BA and another that represents DE, then there are two additional units that represent BE and DA. If there are separate pools of units that represent the thought *New York is dangerous* and the thought *London is safe* then there need to be two additional pools of units to represent the thoughts *New York is safe* and *London is dangerous*. If these conditions are met, then there can, indeed, be a kind of systematicity without compositional structure. But does this approach promise an illuminating account of the systematicity of language understanding or the systematicity of thought?

An advocate of the CMM may argue that this connectionist way of delivering systematicity without compositional structure fails to capture the idea that the capacity to think *New York is dangerous* and *London is safe* is intrinsically linked to the capacity to think *New York is safe* and *London is dangerous*. When the connectionist approach departs from the CMM, it makes systematicity look like an accident. A modeller who adds two separate pools of units to represent the thoughts *New York is safe* and *London is dangerous* could just as well – for all that connectionist representation requires – have added two separate pools of units to represent the thoughts *Life is short* and *Prunes go well with custard*. Connectionists are not without reply, of course, but we shall not pursue the issue further here.

SUMMARY OF SECTION 2

- The computational model of the mind (CMM) is the view that cognition is computation.
- The CMM sees mental processing as involving the rule-guided transformation of structured mental representations.
- Connectionist models are computational models that challenge the idea that mental processing is rule-guided. In connectionist models the primary role is played by the transmission of activation.
- In the past-tense debate, the dual-route words and rules model illustrates the CMM approach. Connectionist modellers have demonstrated that a single-route model can be trained to form the past tense of regular and irregular verbs by transmission of activation.
- Connectionist models do not make use of explicit rules and a network's knowledge of the task domain is embodied in the weights on connections. Nevertheless, there is a sense in which rules can figure in the causal story of how a connectionist model works.
- Connectionist models also challenge the need for mental representations to have compositional structure.

3 MODULARITY

Modularity should have been near the top of the list of concepts and themes relating to theories or models of cognition that you identified in Activity 20.1. But exactly what is meant when a psychological system or function is said to be modular? A notion of modularity is familiar from everyday life. Stereo systems and kitchens have modular components: a CD player connected to an amplifier and a pair of speakers; a 600 mm drawer unit next to a 1200 mm cupboard unit. The electrical wiring in a home is also modular, with one circuit for the lights, another for the sockets, and a separate circuit for kitchen appliances.

In cognitive psychology, it is important that we apply this everyday notion of modular components at the appropriate level, the level of information processing. The brain has anatomical components, but we cannot simply assume that modular information-processing components map directly onto anatomical components of the brain. Cognitive modules are distinguished from each other *functionally*; that is, not by what they *are* neurally, but by what they *do* computationally.

One aspect of the everyday notion of modularity is that modular components are somewhat independent from each other. It is possible to upgrade or even replace an amplifier, while continuing to use the same CD player and speakers, or to replace the CD player with a turntable. Similarly, in the kitchen, the cupboard unit can continue in use even if the drawer unit is removed to make way for a dishwasher. When there is an electrical problem in the home, one circuit may need its fuse replaced (or its circuit breaker reset) while the other circuits continue to operate normally.

David Marr argued that, without modularity, computational processes would be hard to debug or upgrade and he proposed a *principle of modular design* for computational systems: 'a large system can be split up and implemented as a collection of parts that are as nearly independent of one another as the overall task allows' (Marr, 1982, p.102). In a similar spirit, Tim Shallice speaks of modules as 'isolable subsystems' and endorses a definition of modularity proposed by Endel Tulving, who suggested that two psychological systems are functionally different if the following conditions hold:

> *One system can operate independently of the other although not necessarily as efficiently as it could with the support of the other intact system. The operations of one system could be enhanced without a similar effect on the operation of the other; similarly the operations of one system could be suppressed without a comparable effect on the activity of the other. The functional difference also implies that in important, or at least non-negligible ways, the systems operate differently, that is, that their function is governed at least partially by different principles.*
>
> (Tulving, quoted in Shallice, 1988, p.21)

In terms of the everyday notion of modularity, it is relatively uncontroversial that the mind is modular: cognition depends on information-processing systems that have somewhat independent components.
Q1
These modular components are displayed in cognitive psychologists' box-and-arrow diagrams.

Jerry Fodor (1983) has, however, provided us with a more detailed – and more controversial – discussion of modularity in his book *The Modularity of Mind*. This will be the topic for the remainder of this section.

3.1 An outline of Fodor's theory of modularity

According to Fodor's account, the mind is divided into three different types of system: (1) sensory transducers; (2) modular input systems; and (3) non-modular central systems (see Figure 20.3).

The sensory transducers pick up physical stimuli from the environment – photons hitting the retina, sound waves causing the tympanic membrane to vibrate, etc. – and transform these stimuli into a format or code that the brain can understand. In short, transducers produce internal representations in response to stimuli from the external world. Recall that in Chapter 6 (Figure 6.1) you were provided with a picture of the waveform of a segment of speech. That picture is a representation of some of the physical properties of the speech stream. The sensory transducers have to transform those physical properties into a format that the language-processing system can understand. Once external stimuli

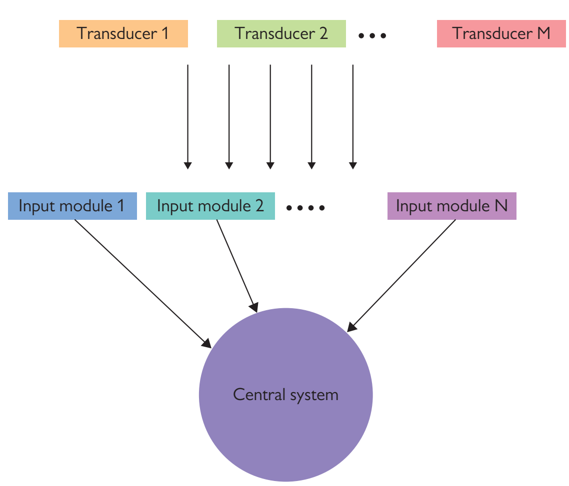

FIGURE 20.3 Fodor's model of perception and cognition. *Source: Coltheart, 1999, Box 1, p.116*

have been internally represented, computation can begin.

The modular input systems mediate between the sensory transducers and the central systems. They deliver, we might say, the perceptual experiences that provide a database of evidence for the processes of belief formation and decision making that are jobs for the non-modular central systems. The non-modular central systems also contain all the encyclopaedic knowledge one has stored in memory. In forming beliefs and making decisions, the central systems will typically take account of evidence from perception and also stored knowledge.

Fodor characterizes the modular input systems as those that possess the following cluster of properties:

- Mandatory operation
- Fast speed of operation
- Fixed neural architecture
- Characteristic and specific patterns of breakdown
- Characteristic pace and sequencing of their ontogeny
- Shallow outputs
- Limited central access to the mental representations that modules compute
- Domain specificity
- Informational encapsulation

This list of nine properties clearly goes well beyond the idea we introduced earlier of a module as a functionally somewhat independent component of a larger information-processing system.

The first five properties on the list are fairly straightforward. When Fodor says (the sixth property) that the final outputs of modules, which are available to central systems, are 'shallow' (1983, p.86), he means that the interface between modular input systems and central systems comes relatively early. For example, when you hear an utterance, the identity of the words and the grammatical structure of the sentence might be computed by a modular input system. But discerning the full significance of the utterance (working out, for example, whether the speaker was asking a question or making a suggestion, being ironic, or intending the utterance to be interpreted metaphorically), and deciding how to respond to the utterance, are tasks for the non-modular central systems. The seventh property of modules – that there is only limited central access to the mental representations that they compute – is the idea that the information at the various intermediate stages of a module's computation is not generally available to central systems.

Fodor does not say that a psychological system must possess *all* of the nine properties on the list if it is to be modular; indeed, he explicitly states that his concept of modularity is a cluster concept and that modularity admits of degrees. However, it is clear from Fodor's other writings that some of these properties – in particular, domain specificity and especially informational encapsulation – are more important to his conceptualization of a modular input system than others (e.g. Fodor, 1985).

3.1.1 Domain specificity

In *The Modularity of Mind*, Fodor's main examples of modular input systems are those responsible for visual perception and for the recognition of spoken language. The claim that these systems are *domain specific* amounts to the idea that these input systems deal only with a limited, idiosyncratic range of stimuli. The language module, for example, only deals with linguistic input. Thus it processes spoken language, but not more general environmental sounds such as the sound of a bell tolling. You should note that, because the language module can also process written language and sign language, the modular input systems do not correspond in a one-to-one way with the five senses.

In cognitive psychology, it is domain specificity that is most often associated with claims about modularity, although the modules proposed are often more fine-grained than those discussed by Fodor. Cognitive psychologists have proposed, for instance, that the visual system includes three modular sub-systems that have the specific jobs of processing faces, processing objects, and processing written words (see Chapters 4 and 6). Moreover, in cognitive psychological models, modules can be nested within one another – for instance, the face-processing module might itself be broken down into sub-modules. Finer-grained domain specificity and modules within modules can be taken as wholly within the spirit of Fodor's account.

3.1.2 Informational encapsulation

Informational encapsulation is the property that is non-negotiable for Fodor as regards whether a system is granted the status of a module. He describes the informational encapsulation of input systems as 'the essence of their modularity' (1983, p.71). We need, therefore, to be clear about what this property involves.

The easiest way to do this is via an example. Look at Figure 20.4. Do the two horizontal lines look the same length? It would be surprising if they did. For most viewers, the top horizontal line *looks* longer than the bottom one. The two lines are, in fact, of equal length (measure them if you want to check). With this knowledge in mind, look at the picture again. Do the two lines *now* look the same length? Once again, it would be surprising if they did. You still see exactly what you did before. But, because you know that the two lines are really the same length, you don't believe what you see. This is an example of *informational encapsulation*. You (now) have some knowledge – about the lines in the Ponzo illusion being the same length – but that knowledge does not affect the processing that takes place within the visual-input system. The visual-input system is, if you like, sealed off from that information or, as Fodor also describes it, visual perception is *cognitively impenetrable* – thoughts and beliefs are unable to penetrate the modular visual input system. The general point – well illustrated by visual illusions such as the Ponzo illusion – is that the processing in a modular input system is not affected by information stored in the central systems.

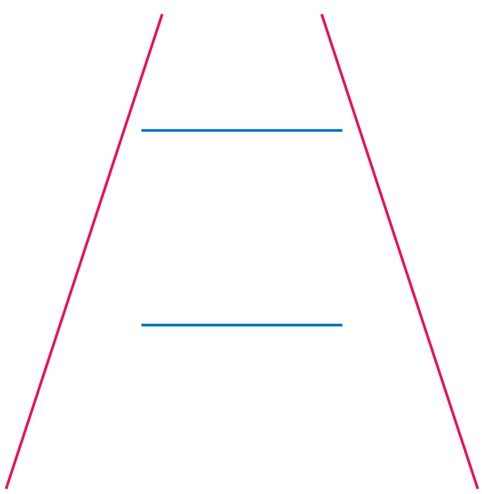

FIGURE 20.4 The Ponzo illusion.

In thinking about modules and the property of informational encapsulation, it is important to observe that Fodor does not rule out top-down information flow *within* a module. The top-down information flow that is excluded or restricted by informational encapsulation is the flow from central systems to modular input systems. The processing in a module does not draw freely on all the information that is stored in the central systems. This point about informational encapsulation has been a source of some confusion about Fodor's views on modularity, so it is worth spending a little more time discussing it.

In *The Modularity of Mind*, Fodor considers an objection to the idea that language comprehension is informationally encapsulated, an objection that stems from our ability to select the relevant meaning of an ambiguous word when it occurs in an appropriate context, such as 'Because Bruce was running short of cash, he went to the *bank*'. One possible explanation of this ability is that contextually relevant knowledge stored in the central systems influences processing in the language-input system, specifically, the process of lexical access. Thus, on the basis of your understanding of the early part of the sentence, your general knowledge about financial matters ensures that only the relevant meaning of the word BANK is accessed; or, more accurately, that only one of the two words spelled BANK is accessed in the lexicon.

Fodor draws on the results of Swinney (1979), which you have already met in Chapter 6, to question this account of contextual disambiguation. We will return to Swinney later, but for now we focus on one aspect of his findings. Swinney used a cross-modal priming paradigm in which participants listened to sentences, some of which contained an ambiguous word such as BUGS. Immediately following this word, a visual word (or non-word) was presented on a screen and the participants' task was simply to make a lexical-decision response ('Is this a word or not?'). In one condition, the ambiguous word BUGS was presented in the context of a sentence that did not bias interpretation towards one or the other of the word's meanings. In this condition, recognition of both the word ANT (semantically related to BUGS = insects) and the word SPY (semantically related to BUGS = secret radio transmitters used in espionage) was facilitated (by comparison with an unrelated word). This finding supports the hypothesis that, when there is no contextual bias, both meanings of BUGS (more accurately, both words with that spelling) are accessed.

In a second condition, the heard sentence strongly biased interpretation of the ambiguous word towards one of its meanings rather than the other. Swinney found that, when BUGS was presented in the context of a sentence about insects, recognition of the word ANT *and* the word SPY was once again facilitated. This finding supports the hypothesis that, when BUGS is first presented, both meanings (more accurately, both words with that spelling) are accessed despite the contextual bias, and that the locus of contextual disambiguation is *after* lexical access.

Fodor suggests that Swinney's findings can be accounted for in terms of the organization of the mental lexicon itself. Perhaps there are links between the lexical entries for semantically related words such that presentation of a word leads to the transmission of activation to lexical entries for semantically related words. In the case of an ambiguous word (e.g. BUGS), there is transmission of activation to lexical entries for words related to both of its meanings (ANT and SPY). Since the lexicon is *internal* to the language-input system, the semantic priming effects would not provide evidence that information from the central systems influences processing in the language-input system. As Fodor warns, 'it makes a difference … where the information comes from' (Fodor, 1983, p.80). If the information does not come from outside the module, it fails to be a counter-example to encapsulation.

✱ ACTIVITY 20.3

Fodor divides the mind into modular input systems and non-modular central systems. Which of the following do you think are modular input systems and which are non-modular central systems? In each case give reasons for your choice.

- Object recognition
- Face recognition
- Spoken word recognition
- Autobiographical memory
- Encoding and retrieval systems
- Problem solving
- Reasoning
- Spoken language production
- Reading aloud
- Attention
- Working memory

COMMENT

This is not as straightforward as it might seem. The earlier stages of object recognition and face recognition certainly seem to be input systems. They are modular in that they are domain specific, but do you agree with Fodor that they are informationally encapsulated? Problem solving seems to be a good candidate for a non-modular central system. However, many would argue that memory is modular in that there seem to be different systems for different kinds of memory (declarative/procedural/autobiographical) – but memory isn't an input system. How should attention be classified? Is it an input system? Is it modular? And what is to be said about output systems? Fodor didn't talk about output systems in *The Modularity of Mind*, but they are often taken to be modular. Do you agree? Consider, for example, one output function – reading a word aloud: which of the criteria for modularity does it meet?

3.2 The central systems

The central systems store information corresponding to knowledge and long-term beliefs and use this information in processes of reasoning, decision making, and planning. According to Fodor, there is an asymmetric flow of information between the modular input systems and the central systems. The input systems pass information – the outputs of modules, the results of their computational processes – to the central systems. But the processing in input systems is encapsulated from information stored in the central systems. This asymmetric flow of information brings clear benefits. If the processing in input systems were to be overly influenced by information already in the central systems – what we already know and believe – then we could never learn anything new from experience.

When we learn from experience, new information from input systems is accepted and old information stored in the central systems has to be revised. But there is a balance to be struck between new information and old information. It would be a very unwise creature that always accepted new information (or misinformation) delivered by an input system and ignored or discarded things that it already knew and believed. You should not always believe your eyes! The balanced integration of new information delivered by input systems with previously stored knowledge and beliefs is a crucial function of the central systems. It leads to the formation of new beliefs and the revision of old ones, and these new or revised beliefs are then used in reasoning, decision making, and planning.

Fodor makes two important claims about properties of the central systems. First, central systems are *domain general* rather than domain specific (they must be able to cope with the outputs from many different input systems). The reason is that it would be very unwise to update beliefs on the basis of the output from just one input system while other input systems were delivering different, and perhaps conflicting, information. Second, central systems are *unencapsulated* – in settling on what to believe, or in making a decision about what to do, anything and everything that the organism knows might, in principle, be relevant.

3.2.1 Argument from analogy: belief formation and hypothesis confirmation

Fodor's argument for the claim that the central systems are unencapsulated depends on an analogy between the processes in the central systems that lead to the *formation of beliefs* and the type of non-demonstrative inference that is characteristic of the *confirmation of hypotheses* in science. This is abductive inference or inference to the best explanation.

Scientific confirmation, Fodor says, is characterized by two properties that he dubs isotropic and Quinean (the latter after the famous twentieth-century American philosopher Willard van Orman Quine, 1908–2000, who argued that our knowledge was organized into a holistic web of belief). Fodor explains these notions as follows:

> *By saying that confirmation is isotropic, I mean that the facts relevant to confirmation of a scientific hypothesis may be drawn from anywhere in the field of previously established empirical … truths. Crudely: everything that the scientist knows is, in principle, relevant to determining what else he ought to believe. In principle, our botany constrains our astronomy, if only we could think of ways to make them connect.*
>
> *By saying that scientific confirmation is Quinean, I mean that the degree of confirmation assigned to a given hypothesis is sensitive to the entire belief system; as it were, the shape of our whole science bears on the epistemic status of each scientific hypothesis.*
>
> (Fodor, 1983, p.105)

These two properties, and the difference between them, need a little spelling out.

Suppose that you are considering two competing empirical hypotheses – perhaps two hypotheses about the generation of past-tense forms of regular and irregular verbs. As a cognitive psychologist, you can devise some behavioural experiments, expecting that the results may support one hypothesis or else the other. But you cannot claim that the evidence from your experiments is the only evidence that can be relevant to the choice between the competing hypotheses. Nor can you claim that the *kind* of evidence yielded by your experiments (perhaps reaction times for a task of producing the past-tense form of a verb) is the only kind of evidence that can be relevant. There is simply no way of putting a boundary around the kinds of evidence that may be relevant to the confirmation or disconfirmation of a hypothesis. As Fodor says of a closely related example: '*Any* facts about the use of language, or about how it is learned, or about the neurology of speaker/hearers, or, for that matter, about the weather on Mars, could, in principle, be relevant to the choice between competing linguistic theories' (1981, p.199). This is what Fodor means when he says that confirmation of hypotheses in science is isotropic. (Fodor also points out (1983, pp.106–7) that the processes of scientific *discovery* – the formation, rather than confirmation, of hypotheses and theories – seem to depend on analogical reasoning and are even more clearly isotropic.)

Saying that confirmation is Quinean is different from saying that confirmation is isotropic, even though being isotropic may sound rather similar to being 'sensitive to the entire belief system'. Suppose that we have two competing empirical hypotheses, or two competing theories, that fit all the relevant evidence equally well. Then we may still choose one theory over the other because it provides a *better explanation*. One theory may have the explanatory virtue of being simpler, or a better fit with the rest of science, or a less radical departure from previous theories (more conservative). In fact, whenever we evaluate competing scientific theories, we consider the explanatoriness of the theories as well as their probability given the available evidence. Fodor points out that, when theories are assessed for their explanatoriness, they are considered, not in isolation, but against the background of the rest of science. This is particularly clear in the case of the explanatory virtue of offering a good fit with the rest of science (offering good prospects for unification of phenomena that are superficially disparate). But the same point holds for the explanatory virtues of simplicity and conservativeness as well. Thus, the choice between competing hypotheses, or competing theories, depends on 'the shape of our whole science' (Fodor, 1983, p.105). This is what Fodor means when he says that confirmation of hypotheses in science is Quinean.

If Fodor is right to propose that the confirmation of hypotheses in science is a model for the formation of beliefs in the central systems, then we should expect that the operation of the central systems is isotropic and Quinean. An example illustrating what it means for central system processes of belief formation to be isotropic and Quinean is provided in Box 20.1. If the operation of the central systems *is* isotropic and Quinean then the central systems are not modular and, in particular, are not encapsulated. Processes of belief formation need to have free access to anything and everything that we know and believe.

This is Fodor's argument from analogy, but is there any direct empirical evidence that the central systems are unencapsulated? Is scientific confirmation really a good model of ordinary belief formation? Unfortunately, according to Fodor, evidence on this is scanty 'given the underdeveloped state of psychological theories of thought and problem solving' (Fodor, 1983, p.112).

3.2.2 Unencapsulated central systems and the CMM

As by his lights evidence is lacking, Fodor points to a difficulty in the development of theories of the central systems that is 'just the sort [of problem] we should expect to encounter if such processes are, in essential respects, Quinean/isotropic rather than encapsulated' (1983, p.112). It concerns the difficulties posed for artificial intelligence by the so-called frame problem. This is, very roughly, the problem of how to build a robot that can update its beliefs about the world as a result of the action it takes (Dennett, 1984). Consider, for example, a robot that is given the job of making a phone call to Mary:

> *Let's assume the robot 'knows' it can get Mary's number by consulting the directory. So it looks up Mary's number and proceeds to dial. So far, so good. But now, notice that commencing to dial has all sorts of direct and indirect effects on the state of the world (including, of course, the internal state of the robot), and some of these effects are ones the device needs to keep in mind for the guidance of its future actions*

BOX 20.1 Insects or espionage?

Suppose that you overhear a man in a dark suit utter an isolated sentence in which the word BUGS occurs. Perhaps he is talking about insects, or perhaps about secret radio transmitters. Suppose that, for some reason, it is crucial that you should reach the correct belief as to whether the man is talking about insects or espionage. Then, anything and everything that you know might be relevant.

First, *any* piece of evidence might turn out to be relevant. Does the man look as if he works for a pest control company, or does he look as if he might be working for (or against) an intelligence agency? Was there something in the newspaper about an invasion of insect pests, or about a spy cell being active? What about that report on the web about an interdisciplinary team of astronomers and biologists who claimed to have discovered a correlation – as yet unexplained – between the weather on Mars and the reproduction of ants? If there is a correlation, then the weather on Mars is relevant, just as Fodor said. But what about that other team who said they had discovered a different correlation between the weather on Mars and increased international tensions? Which team of scientists is to be trusted? Which universities are they from, and how good is the research there? So university research rankings are relevant as well. And so on, without end. If you really need to know whether it is more likely that the dark-suited man was talking about insects or about espionage, there is no good way of putting a boundary around the potentially relevant considerations and ignoring the rest of what you know. Belief formation is isotropic.

Second, the overall shape of your *entire* body of knowledge is important as well. Which hypothesis would fit better with what you already know and believe? How much revision to your other beliefs would be needed if you were to adopt the belief that the man was talking about insects because there is an infestation of ants, or if you were to believe that he was talking about espionage because there are secret radio transmitters everywhere? Which hypothesis would, perhaps after some initial upheaval in your other beliefs, allow the simpler and more coherent total world view? Belief formation is Quinean; it is 'sensitive to global properties of belief systems' (Fodor, 1983, p.111).

and expectations. For example, when the dialing commences, the phone ceases to be free to outside calls; the robot's fingers (or whatever) undergo appropriate alterations of spatial location; the dial tone cuts off and gets replaced by beeps; . . . and so forth. Some (but, in principle, not all) such consequences are ones the robot must be designed to monitor since they are relevant to 'updating' beliefs upon which it may come to act. Well, which *consequences?*

(Fodor, 1983, p.113)

We, of course, do this everyday updating effortlessly. But, at least on Fodor's reading of the literature, AI researchers have found building a device that can do it intractable.

It may be possible for AI researchers to solve or sidestep the frame problem, either by adopting ambitions more modest than building a robot with human intelligence, or by abandoning the computational paradigm that is the basis of the CMM. But Fodor regards the frame problem in AI as illustrating a much more

✱ ACTIVITY 20.4

Reconsider Chapters 10 and 12 on problem solving and on reasoning. Is there any evidence provided in those chapters that would either support or undermine Fodor's views on the central systems? Is he right that psychological theories of thought and problem solving are underdeveloped?

general problem about the processes of updating or revising beliefs. The problem arises if the processes involve updating in the light of the consequences of one's own actions (as in the example of the robot making a phone call). But it also arises if the updating is based on new information of other kinds (delivered, for example, by a perceptual input system). The processes of belief formation and belief revision need to be isotropic. There is no *a priori* method for putting a boundary around the considerations that might be relevant. And, as in AI, so more generally, this property of processing in the central systems presents a challenge to computational theories and, specifically, a challenge to the CMM (Shanahan, 2009).

Fodor believes that the prospects for a computational psychology of the central systems are bleak. Indeed, he proposes a principle that, he suggests, 'some day will come to be known as "Fodor's First Law of the Nonexistence of Cognitive Science"'; namely, that 'the more global (e.g. the more isotropic) a cognitive process is, the less anybody understands it. *Very* global processes, like analogical reasoning, aren't understood at all' (Fodor, 1983, p.107). According to Fodor, the CMM just will not explain the central systems (Fodor, 2000 gives a book-length treatment of this issue).

3.3 Debates about modularity

Fodor's account of modularity has a positive part – input systems are modular – and a negative part – central systems are not modular. The account has been very influential, but also controversial. Many of the criticisms of the positive part of Fodor's account focus on the property of informational encapsulation, which Fodor regards as the essence of modularity.

One important kind of challenge takes the form of presenting evidence that a system that Fodor regards as an input system is not really informationally encapsulated. A defender of modularity might respond by saying that the processing that is influenced by background knowledge and beliefs takes place in the central systems, after the encapsulated input system has delivered its output; the locus of top-down influence is post-perceptual. This response would fit well with the sixth property of modules, namely that their final outputs are *shallow*. But here we must be careful because Fodor actually introduced a further property of input systems, one that we have not previously mentioned. Fodor regards the input systems as delivering information to the central systems so that it can be used in *conscious* inferences. So the response would also need to take into account the idea that the outputs of modules are accessible to consciousness or 'phenomenologically salient' (Fodor, 1983, p.87).

Suppose, for example, that it seemed as though visual object recognition was not informationally encapsulated. A response would be to say that visual object recognition takes place in the central systems and not in the modular visual-input system. In Marr's theory of visual processing (Chapter 3), the primal sketch and the 2½D sketch are shallow levels of representation and, to that extent, they would appear to be suitable candidates for the output of a visual-input system. But for Fodor they do not meet the requirement of phenomenological salience – 'from the point of view of phenomenological accessibility, perception is above all the recognition of objects and events' (Fodor, 1983, p.94) – and so the response would fail.

According to Fodor's account of the mind, the processing in input systems is informationally encapsulated *and* the outputs of input systems are accessible to consciousness (phenomenologically salient). These two requirements may sometimes impose conflicting pressures on the location of the interfaces between input systems and central systems.

3.3.1 Arguments against informational encapsulation

According to Fodor, in the processing of spoken language the detection of irony and the interpretation of metaphor are unlikely to be informationally encapsulated. Those processes belong in the central systems because they require access to our beliefs, including beliefs about the speaker's intentions, about whether what the speaker said would make sense if interpreted literally, and much else. So, at what point, earlier than those processes of pragmatic interpretation, is the interface between the language-input system and the central systems located? Fodor's proposal is that the language-input system delivers information about the linguistic form of the sentence that was uttered – which words it contains and how they are put together. This output is relatively shallow – it certainly falls well short of complete understanding of the message that the speaker was intending to get across – but it is still phenomenologically salient.

In a critique of Fodor's account of modularity, Marslen-Wilson and Tyler (1987) review a range of evidence that seems to show that the language-input system, as conceived by Fodor, is not informationally encapsulated. Processing in the language-input system is, they argue, influenced by knowledge about the non-linguistic world. In one experiment (Marslen-Wilson *et al.*, 1988), participants listened to sentences, some of which were anomalous in various ways. For our purposes, the crucial comparison is between sentences in normal prose, such as 'The boy held the guitar', and sentences containing a pragmatic anomaly, such as 'The boy buried the guitar'. The participants' task was simply to press a button as soon as they detected a previously specified target word – in the case of these examples, the word GUITAR. The results show that participants responded significantly faster when the target word was presented in the normal sentence than when it was presented in the pragmatically anomalous sentence.

This finding is important because the anomalousness of the second sentence does not result from the syntactic properties of BURY and GUITAR, and it is not explained by the literal meanings of those words, as specified within the language-input system. As the authors explain: 'The pragmatic oddness of burying a guitar is something that we have to infer, *given our knowledge of the world*, and given what we know about guitars, the likely effects of burying them, and so on' (Marslen-Wilson *et al.*, 1988, p.6; emphasis added). Thus, the finding provides evidence that knowledge of the non-linguistic world influences on-line processing of spoken language input.

In response to this finding, a defender of modularity might propose that the locus of top-down influence is post-perceptual. In a little more detail, the proposal would have two components. First, it would be argued, spoken word recognition is an informationally encapsulated process that takes place within the language-input module. But second, so the argument would continue, deciding to press the button is not an informationally encapsulated process, and is therefore a matter for the central systems. Even under speeded conditions, the decision depends on the participant's beliefs. A slower response to the word GUITAR when it is presented in the pragmatically anomalous sentence might therefore reflect 'the hearer's inability to believe that the speaker could have said what it sounded like he said' (Fodor, 1985, p.5).

Other experiments reviewed by Marslen-Wilson and Tyler (1987) provide evidence that non-linguistic knowledge influences the resolution of ambiguity in phrases such as 'visiting relatives' or 'landing planes' (Tyler and Marslen-Wilson, 1977) and influences the interpretation of anaphors[2] (Tyler and Marslen-Wilson, 1982; Marslen-Wilson *et al.*, 1993). In each case where informational encapsulation is challenged, the defender of modularity may respond by claiming either (i) that the top-down influence is coming from within the input system and not from the central systems, or else (ii) that the processing that is being influenced occurs in the central systems and not within the input system. And, of course, in each case there can be further debate about whether the response is adequate.

Let us now return to Swinney's (1979) experiment, and the question as to whether *lexical access* is informationally encapsulated. Swinney found that both ANT and SPY are primed when they are presented immediately following BUGS and, as we have seen, these semantic priming effects could be explained in terms of the organization of the mental lexicon, and not in terms of information from the central systems. However, Swinney also discovered that if BUGS is heard in a sentence about insects and ANT or SPY is presented three syllables (about one

[2] Anaphors (which may be anaphoric expressions such as pronouns or may be anaphoric devices that are not phonologically explicit in spoken language) refer back to someone or something that has been introduced or identified by an antecedent in the previous text or linguistic context.

second) after BUGS, then ANT is primed but SPY is not. This indicates that, while both meanings of BUGS (both words spelled BUGS) are accessed initially, there is then a disambiguation process in which one of them is selected at the expense of the other. It is plausible that this process of selection draws on information from the central systems, but the locus of top-down influence can be argued to occur *after* lexical access. Indeed, Swinney (1979, p.657) suggests that 'a very rapid postaccess decision process is at work'.

Marslen-Wilson and Tyler (1987) also raise some more general issues for Fodor's account of modularity. Fodor connects fast processing in input systems with the properties of being mandatory and being informationally encapsulated. But the influence of non-linguistic knowledge does not make processing slow. In the experiment that we described earlier (Marslen-Wilson *et al.*, 1988), the difference in mean reaction times to target words between the normal and pragmatically anomalous sentences was just 28 msec. So processing that involves central systems can still be fast and speed of processing does not allow us to draw a clear boundary around the modular language-input system. Perhaps surprisingly, it is also not straightforward to use the notion of informational encapsulation – the essence of modularity – to draw the boundary between an input system and the central systems. The reason is that the property of informational encapsulation is defined in terms of the absence of top-down influence from information that is stored *in the central systems*.

Fodor says that each module has a 'proprietary database' of information that the module's computations can draw on (Fodor, 1985, p.3). Informational encapsulation requires only that this data should not include all the information that is stored in the central systems. Domain specificity suggests, in addition, that the proprietary database of the language module, for example, should include only information about language – information about phonology and syntax, and information about words (a lexicon), but not information about whether it is likely that someone would bury a guitar. Against this suggestion, Marslen-Wilson and Tyler (1987) urge that we should not so readily assume that *language processing* follows the contours of *linguistic theory*. More generally, there is still work to be done before we can confidently, and non-arbitrarily, draw boundaries between input systems and the central systems.

3.3.2 Prospects for the psychology of the central systems

The negative part of Fodor's account of modularity is that central systems are not modular and, as we have seen, Fodor is pessimistic about the prospects for a computational psychology of the central systems. Critics of this negative part of the account argue that the mind is modular throughout – *massively modular* (Sperber, 1994, 1996, 2002) – and that, consequently, central cognitive processes can also be brought within the scope of the CMM.

The massive modularity hypothesis draws support from evolutionary psychology (Barkow *et al.*, 1992; Cosmides and Tooby, 1994). The idea is not that there is a module for thinking, another for belief formation, and another for decision making, but rather that there is a host of more specialized modules, each dedicated to the solution of a particular, and perhaps quite

✱ ACTIVITY 20.5

The Marslen-Wilson and Tyler data concern the language-input module. Can you find any evidence – from this book or elsewhere – that is relevant to Fodor's claim that the *visual*-input system is informationally encapsulated?

COMMENT

You might start by looking again at Chapter 3, Section 5 on 'Constructivist approaches to perception'. That section describes evidence supporting the hypothesis that knowledge affects what we perceive. How might Fodor respond?

idiosyncratic, type of problem. Dan Sperber argues that the massive modularity hypothesis:

> *gets support from a wealth of recent work . . . tending to show that many basic conceptual thought processes found in every culture and in every fully developed human are governed by domain-specific competences. For instance it is argued that people's ordinary understanding of an inert solid object, of the appearance of an organism, or of the actions of a person are based on three distinct mechanisms: a naive physics, a naive biology, and a naive psychology.*
>
> (Sperber, 1996, p.123)

In a similar vein, Steven Pinker says, at the beginning of his book, *How the Mind Works*: 'the mind is a system of organs of computation designed by natural selection to solve the problems faced by our evolutionary ancestors in their foraging way of life' (Pinker, 1997, p.x). You may not be surprised to learn that Fodor argues against massive modularity in his book, *The Mind Doesn't Work That Way* (2000; for further discussion, see Pinker, 2005a; Fodor, 2005; Pinker, 2005b).

If massive modularity is to provide grounds for optimism about the prospects for scientific understanding of central cognitive processes then one thing that needs to be demonstrated is how processes such as belief formation, decision making, and analogical reasoning could be carried out by a massively modular cognitive architecture. Ideally, it would be shown how those processes could, after all, fall within the scope of the CMM. A sophisticated defence of massive modularity that does attempt to show these things is provided by Peter Carruthers (2006).

As you consider this debate about the negative part of Fodor's account of modularity, you should be aware that complications could arise from the fact that advocates of massive modularity may not mean the same by the term 'module' as Fodor means. A naïve physics, or biology, or psychology, as described by Sperber, sounds like a body of knowledge. But Fodor's modules are processing systems, rather than stored bodies of domain-specific knowledge. Carruthers relaxes Fodor's account of modules and, in particular, does not require that modules are informationally encapsulated – the essence of Fodorian modularity. So it seems that an advocate of massive modularity might actually accept the negative part of Fodor's account – the central systems are not modular in Fodor's sense – while maintaining that central cognitive processes are underpinned by systems that are modular in a different way.

You would then need to determine where, if anywhere, the real disagreement between the two sides is located (Robbins, 2009). Do the advocates of massive modularity accept that cognitive processes of belief formation have the properties of being isotropic and Quinean? Do they accept that those properties require that the processes are unencapsulated? Do they accept that processes that are unencapsulated are not computationally tractable?

SUMMARY OF SECTION 3

- Fodor proposes that the mind can be divided into sensory transducers, modular input systems, and non-modular central systems.
- Modular input systems are characterized by a cluster of properties, but the properties of being domain specific and informationally encapsulated are especially important.
- Informational encapsulation means that the processing undertaken by a module is not affected by knowledge stored in the central systems. Top-down information flow within a module is permitted.
- The central systems receive output from the modular input systems and are involved in belief formation and decision making. They are domain general and unencapsulated.
- Fodor proposes that a good model for the functioning of the central systems is the kind of non-demonstrative inference that is used in scientific confirmation – that is, abductive inference

or inference to the best explanation. This kind of inference has the properties of being isotropic and Quinean.

- Fodor's 'First Law of the Non-existence of Cognitive Science' warns that the prospects for a computational psychology of the central systems are bleak.
- Fodor's account of modularity has a positive part – input systems are modular – and a negative part – central systems are not modular.
  - Critics of the positive part of Fodor's account of modularity have questioned whether input systems are really informationally encapsulated.
  - Critics of the negative part of Fodor's account have proposed that the mind is modular throughout.

4 COGNITIVE NEUROPSYCHOLOGY

We now turn to a branch of cognitive psychology that relies on an assumption of modularity to draw conclusions about cognitive processes. Cognitive neuropsychology uses data from patients with acquired cognitive disorders following brain injury in order to constrain models and theories of normal cognition (Chapter 13). The term 'neuropsychology' is also used for a discipline that is concerned with the relationship between brain and behaviour (and is thus close to behavioural neurology). Cognitive neuropsychology and neuropsychology in this other sense have common antecedents in nineteenth-century neurology. But cognitive neuropsychology is different from neuropsychology in being an approach or research programme within cognitive psychology. Its subject matter is not the brain, but the information-processing structure of the mind.

For cognitive neuropsychology, the most important pattern of impaired cognitive functioning in patients is double dissociation of impairments on two cognitive tasks: one patient (A) shows impaired performance on Task 1 but performs at a normal or near normal level on Task 2, while a second patient (B) shows the reverse pattern, impaired on Task 2 but unimpaired or near normal on Task 1. From this evidence, it is inferred that the two tasks involve separate information-processing routes or mechanisms.

In order to draw these inferences about the structure of normal cognition, cognitive neuropsychology relies on several assumptions:

1. The first of these is an assumption of *modularity*: the normal cognitive system is made up of somewhat independent components or information-processing mechanisms. However, the notion of modularity that is used in the practice of cognitive neuropsychology has fewer commitments than Fodorian modularity. Two points of contact between the modularity assumption in cognitive neuropsychology and issues in Section 3 of this chapter are briefly described in Box 20.2 and Box 20.3.
2. The second assumption is *universality*: the modular structure or *functional architecture* of the mind is the same, in relevant respects, for all normal (neurologically healthy) individuals (Caramazza, 1986).
3. The third assumption is that, when one cognitive component is damaged, this does not bring about substantial reorganization of the prior modular structure. The undamaged components continue to operate as before and no new components are added. This assumption is sometimes called *transparency* (Caramazza, 1986) and sometimes *subtractivity* (Coltheart, 2001). (See Chapter 13 for further discussion.)

BOX 20.2 Modularity in central systems

Fodor's main argument for the negative part of his account of modularity was reviewed in Section 3.2. The central systems are not modular because central cognitive processes, such as belief formation, have properties – being isotropic and Quinean – that require them not to be informationally encapsulated. But, if the notion of modularity that is used in cognitive neuropsychology is closer to the everyday notion than to Fodor's, then there may be modules – in the sense of somewhat independent components – in the central systems. Damage to separate components of the central systems might, in principle, result in selective cognitive impairments and, indeed, there is evidence that memory systems can be selectively impaired, that the ability to do arithmetical calculation can be destroyed, that attentional systems can be damaged, and that everyday psychological understanding (theory of mind) can be impaired (Shallice, 1988; Ellis and Young, 1998; and Baron-Cohen *et al.*, 1999 describe evidence of all these impairments).

One of the properties associated with Fodorian modularity is fixed neural architecture and Fodor suggests that this property, like informational encapsulation, is missing from the central systems and that, perhaps, 'central problem-solving is subserved by equipotential neural mechanisms' (Fodor, 1983, p.119). Here, Fodor alludes to Karl Lashley (1930), who maintained that some functions (such as a rat's capacity to run through a maze) are localized in the whole cortex and can be impaired by damage anywhere in the cortex. If several separate components of the central systems were each to be localized in the whole cortex then it would be difficult for them to be separately damaged and selective impairments would not be expected. On the face of it, Fodor's suggestion that the neural substrate of central cognitive functions is diffuse may be in tension with the evidence of selective central cognitive impairments. (One possibility to consider would be that the neural basis of central cognitive *processing* is diffuse but that domain-specific *knowledge modules* can be separately damaged.)

There is also a more general issue here. Cognitive neuropsychology is a division of cognitive psychology and its subject matter is the information-processing structure of the mind. But the practice of cognitive neuropsychology depends on an additional assumption about the brain – an assumption of (at least some) *anatomical modularity*. If selective impairments are to be explained in terms of damage to components of the mind – components of input systems or of the central systems – then components that are functionally separate (modules) must be able to be damaged separately. This would not be possible if many separate modules all shared the same neural substrate.

4.1 Inference in cognitive neuropsychology

In cognitive neuropsychology, inferences about the modular structure of the normal cognitive system are drawn from patterns of cognitive impairment in patients, and the most important pattern is double dissociation. How do those inferences work and why is double dissociation regarded as evidence for separate cognitive modules? We can answer this question in three steps.

First, *why is evidence from brain-damaged people relevant to models of normal functioning at all?* There is no *a priori* way of putting a boundary around the kinds of evidence that may be relevant to the confirmation or disconfirmation of a hypothesis (see our earlier discussion of hypothesis confirmation as being isotropic). So there is no *a priori* reason to suppose that evidence from brain-damaged people will or will not be relevant to models of normal cognitive functioning. This is an empirical issue. However, if the *modularity* assumption about the normal cognitive system is correct then there may be separate

BOX 20.3 A non-Fodorian definition of modularity

Coltheart (1999) proposes that a module should be defined as 'a cognitive system whose application is domain specific' (1999, p.118) and that it should be left as an empirical question whether a particular domain-specific module has any of the other properties of Fodorian modules, such as having fixed neural architecture or being informationally encapsulated.

The notion of a module as a domain-specific processing system may well be attractive to cognitive psychologists who believe that input systems, such as the language-input system, are not informationally encapsulated. Also, Coltheart's notion has an advantage of generality, in that it may apply to some central cognitive systems as well as to input systems. But, if the notion is to be put to use, then it requires a good account of what domain specificity is – and what a domain is. Fodor himself warns that we have to be careful to avoid a 'trivial kind of domain specificity' (Fodor, 1983, p.48). Someone might say that cows and sheep are two different domains and that cow perception and sheep perception are domain-specific processes. But this would be a purely verbal manoeuvre if the same mechanisms of visual perception are involved in perceiving both cows and sheep.

Coltheart recognizes the need to provide an account of domain specificity that is empirically based, rather than merely verbal. So he invites us to imagine that we started out with the idea that there is a single visual-recognition system for faces, objects, and written words: 'Then we noticed that in the neuropsychological literature there were reports of patients with impaired visual word recognition but who retained face recognition and of patients with impaired visual word recognition but who retained visual object recognition. We also noticed reports of patients with impaired visual object recognition but who retained face recognition or visual word recognition . . . This collection of results refutes our original idea . . . Instead, it suggests that there are three separate modules' (Coltheart, 1999, p.119).

This is a typical example of inference in cognitive neuropsychology. Case studies of patients reveal dissociation of impairments between tasks – in this case, multiple dissociations amongst the three tasks of recognizing faces, objects, and written words. This evidence supports the hypothesis that the three tasks involve three separate information-processing systems, each of which is domain specific. Faces, objects, and written words – unlike cows and sheep – are empirically (and not just verbally) different domains. Thus, Coltheart's notion of a module as a domain-specific processing system seems to fit well with the practice of cognitive neuropsychology.

cognitive modules involved in different cognitive tasks. If the assumption of *anatomical modularity* is also correct, then the separate cognitive modules may be independently damaged. If the subtractivity/transparency assumption is correct, then damage to one module will result in impairment of functions that involve that module, while other functions are unimpaired (or spared). So the cognitive neuropsychology framework has the potential to explain patterns of selective impairment in patients following brain injury. If a hypothesis about separate modules in the normal cognitive system allows the *best explanation* of a pattern of impairment then the hypothesis is, to that extent, supported – by inference to the best explanation.

It might be, of course, that these assumptions are incorrect. Brain injury might, in principle, have resulted in apparently random patterns of impairment that could not be understood at all in terms of models of normal cognition. But it is a consistent neuropsychological finding, from the nineteenth century until the present, that brain injury often results in selective impairment of specific cognitive functions.

Second, *why does double dissociation provide better evidence for separate modules than a single (one-way) dissociation?* If the *universality* assumption is correct, then the modular structure of the normal cognitive system is the same in all neurologically healthy individuals. Suppose that, in healthy individuals, two separate modules, M1 and M2, are involved in two different tasks, Task 1 and Task 2. Then independent damage to each of the modules in two patients following brain injury provides a good explanation of double dissociation of impairments on the two tasks. This good explanation may turn out to be the best explanation.

Suppose that there were only a one-way dissociation: a patient shows impaired performance on Task 1 but performs at a normal level on Task 2. One possible explanation would be that two separate modules, M1 and M2, are involved in the two tasks and that, in this patient, M1 is damaged while M2 is intact. But there is also a possible alternative explanation. It might be that just the same modules are involved in the two tasks but that Task 1 is more difficult – more demanding of cognitive resources – than Task 2. If that were so then, after partial damage to the modules that are involved in both tasks, performance on Task 1 might be impaired while performance on Task 2 remained normal.

A double dissociation, in which a second patient shows the reverse pattern (normal performance on Task 1 but impaired performance on Task 2), is better evidence for separate modules than the single dissociation because it excludes the possible alternative explanation in terms of the greater difficulty of one of the tasks. But, of course, excluding one possible alternative explanation is very far from providing a logical guarantee that the hypothesis of separate modules involved in the two tasks is correct. Normal science proceeds by inference to the best explanation and evidence does not provide logical guarantees but, at best, a change in the balance of probabilities between competing theories or models (Coltheart and Davies, 2003; Davies, 2010).

Third, just as impairments can be dissociated, they can also be associated. *If impairments on two tasks usually occur together, can we infer that just the same modules are involved in the two tasks?* The hypothesis that, in the normal cognitive system, the same modules are involved in Task 1 and Task 2 certainly provides one possible explanation of the co-occurrence of impairments on the two tasks following brain injury. But there is a fairly obvious possible alternative explanation in terms of neuroanatomy. Even if functionally separate modules, M1 and M2, were involved in the two tasks, it might be that brain injury that damaged one of these modules would inevitably damage the other one as well. The two modules might be localized in the same small region of the brain, or their positions relative to a blood vessel in the brain might make it overwhelmingly likely that a blockage in or bleeding from that blood vessel would damage both modules. In short, there might not be sufficient anatomical modularity to allow the two modules to be damaged independently.

4.2 Cognitive neuropsychology and the past-tense debate

In this section we return to the past-tense debate (Section 2.3) and consider how neuropsychological evidence has contributed to the debate.

4.2.1 Double dissociation evidence

Evidence of double dissociation on tasks related to the past tense of regular and irregular verbs would support the hypothesis that separate modules are involved in the formation of the past tense of regular and irregular verbs. It would favour dual-route models, such as the words and rules model, over single-route connectionist models (e.g. Rumelhart and McClelland, 1986; Plunkett and Marchman, 1993).

Double dissociations were described in two papers published in 1997. Ullman and colleagues (1997) used an experimental paradigm in which participants read aloud pairs of sentences, such as, 'Every day I *dig* a hole. Just like every day, yesterday I ________ a hole', and filled in the blank in the second sentence. Patients suffering from Alzheimer's disease with severely impaired memory for words performed the task significantly worse (by comparison with control participants) for irregular verbs (e.g. DIG) than for regular verbs (e.g. LOOK) or novel verbs (e.g. PLAG). Patients with Parkinson's disease and severely suppressed movement (hypokinesia) showed the reverse pattern, performing significantly worse (by comparison with control participants) for regular verbs, and much worse for novel verbs, than for irregular verbs. Ullman and colleagues argue that the double dissociation supports 'psycholinguistic theories that emphasize grammar and lexicon as distinct components … especially in the treatment of regular and irregular grammatical phenomena' (1997, p.274; see also Ullman *et al.*, 2005).

Marslen-Wilson and Tyler (1997, 1998) used an auditory priming paradigm that did not require participants to read text or give spoken responses. Participants listened to a spoken target word, immediately preceded by a prime. The prime was either morphologically related to the target (JUMPED–JUMP; GAVE–GIVE), semantically related to the target (GOOSE–SWAN), or unrelated to the target. Filler stimuli included non-words, and the participants' task was simply to make a lexical-decision response ('Is this a word or not?'). The crucial comparison was between priming of regular verbs by their past-tense forms (e.g. JUMPED–JUMP) and priming of irregular verbs by their past-tense forms (e.g. GAVE–GIVE). In this primed lexical decision paradigm, neurologically healthy individuals responded faster to targets when the prime was morphologically or semantically related than when the prime was unrelated, and both regular and irregular verbs were primed by their past-tense forms.

However, three patients did show significant differences in priming between regular verbs and irregular verbs. Two patients, with impaired comprehension and production of inflected forms following left-hemisphere stroke, exhibited priming for irregular verbs but not for regular verbs. A third patient, with symptoms of Broca's aphasia following damage to both hemispheres, exhibited the reverse pattern: priming for regular verbs but not for irregular. Marslen-Wilson and Tyler argue that this double dissociation suggests 'the two morphological categories ally themselves with different types of mental computation' (1997, p.593).

4.2.2 The connectionist response

This evidence of double dissociation supports dual-route, rather than single-route, models of formation of the past tense. But it certainly does not provide a logical guarantee that the words and rules model is correct. It does not even provide a logical guarantee that some dual-route model is correct and all single-route models are incorrect.

Kim Plunkett and colleagues (Juola and Punkett, 2000; Plunkett and Bandelow, 2006) have made this point vivid by demonstrating that double dissociation of impairments may arise from *random* damage to a *single-route* network. They trained a single-route network to form the past tense of verbs (and the plural of nouns). The network (analogous to a neurologically healthy individual) was then copied more than 13,000 times, and each copy subjected to random damage (by removal of 1 per cent of its connections). Amongst these 'lesioned' networks (analogous to patients following brain injury) there were pairs of networks instantiating double dissociation of impairments between regular and irregular verbs. In short, double dissociation evidence from damaged networks may be compatible with the undamaged network having only a single route.

Joanisse and Seidenberg (1999) constructed a single-route connectionist model of past-tense formation that was different from earlier connectionist models in an important way: it included semantic representations of verbs as well as phonological representations. By subjecting components of the model to damage, Joanisse and Seidenberg were able to simulate the double dissociation that Ullman and colleagues (1997) found in patients suffering from Alzheimer's disease and Parkinson's disease. When the semantic representations in the trained network were damaged, the effect was greatest for irregular verbs (as in the patients with Alzheimer's disease); when the phonological representations were damaged, the effect was greatest for novel verbs (as in the patients with Parkinson's disease). Thus, in opposition to the words and rules model, Joanisse and Seidenberg defend a single-route theory of past-tense formation: 'The same network structure is used in processing all words. …

The explanation for the behavioral dissociations lies with the fact that the network includes distinct phonological and semantic representations' (1999, p.7597; see also Joanisse and Seidenberg, 2005 for a similar account of neuroimaging evidence).

4.2.3 Cognitive neuropsychology and connectionist modelling

Twenty-five years or more after Rumelhart and McClelland published their connectionist model of the English past tense, the debate continues. It may be that the empirical evidence will not adjudicate decisively between the competing accounts of past-tense formation and that dual-route words and rules models and single-route connectionist models will do about equally well in accounting for the data – not only data from behavioural experiments but also neuropsychological and neuroimaging evidence. (For a recent connectionist model and new behavioural evidence, see Woollams *et al.*, 2009.)

There is a parallel debate about reading words aloud. Again, there are competing models – the dual-route DRC model (Coltheart *et al.*, 2001) and the single-route connectionist models (Seidenberg and McClelland, 1989; Plaut *et al.* 1996). Here, too, double dissociation – between surface dyslexia and phonological dyslexia – has played an important role. Surface dyslexia (impaired reading of irregular words while reading of regular words and non-words is relatively spared) is explained in terms of damage to the lexical route in the DRC model, and phonological dyslexia (impaired reading of non-words while reading of regular and irregular words is relatively spared) is explained in terms of damage to the non-lexical route, with its spelling-sound rules. However, a single-route connectionist model (Plaut *et al.*, 1996) explains the double dissociation in terms of damage to semantic representations or to phonological representations.

In both these areas – the past tense and reading aloud – continuing work is providing new evidence against which competing models can be assessed. Aside from asking which kind of model, rule-based or connectionist, can better fit the data, researchers in these areas are asking another question. Which kind of model offers the *better explanation* of the available evidence – which style of explanation has broader scope and offers better prospects for unifying seemingly disparate phenomena, which is more parsimonious, which achieves greater depth, which leads to the more interesting new predictions? These issues about fitting and explaining data are at the forefront of contemporary debate (Coltheart, 2006a; Seidenberg and Plaut, 2006).

SUMMARY OF SECTION 4

- Cognitive neuropsychology uses data from patients with acquired cognitive disorders following brain injury in order to constrain models and theories of normal cognition.
- For cognitive neuropsychology, the most important pattern of impaired cognitive functioning in patients is double dissociation of impairments on two cognitive tasks.
- Neuropsychological evidence of double dissociation of impairments has contributed to the past-tense debate because it appears to support dual-route models over single-route models.
- Connectionist modellers have demonstrated that double dissociation of impairments may result from random damage to a single-route model.
- Double dissociation of impairments may also result from damage to separate semantic and phonological representations in a connectionist network, even though regular and irregular verbs are processed through the same connections.
- In the past-tense debate, and in a parallel debate about reading aloud, researchers ask not only which kind of model can fit the data better, but also which offers the better explanation of the available evidence.

5 COGNITIVE PSYCHOLOGY AND THE BRAIN

In this final section, we turn to one aspect of the relationship between mind and brain – a large topic with a very long history reaching back at least to ancient Greek philosophy, more than 2,000 years ago. The specific aspect that we shall discuss is the relationship between cognitive psychology and the study of the brain – that is, neuroanatomy or the study of neural circuits (we shall call this neurobiology in what follows).

5.1 Levels of explanation

If you look back over the previous chapters in this book, you will find relatively little discussion of the way in which cognitive systems are actually implemented by the brain. Why is this? There are, we suggest, two related reasons. The first is the influence of David Marr's (1982) account of the three levels at which an information-processing system can be understood. As you saw in Chapter 1, these three levels are:

- Level 1: The level of computational theory, where we ask what is being computed, and why it is being computed. When we know what is being computed, we can also assess how, in principle, it could be computed.
- Level 2: The level of representation and algorithm, where we ask how the computation that is abstractly described at level 1 is actually carried out in the system that is under investigation. More specifically, we ask in what formats input information and output information are represented, and by what algorithm the transformation from input to output is accomplished.
- Level 3: The level of hardware implementation, where we ask how the representations and algorithms specified at level 2 are physically realized. When we are trying to understand information processing in human cognitive systems, we ask how the representations and algorithms specified at level 2 are physically realized in the human brain.

Experimental research in cognitive psychology is generally aimed at understanding actual representational structures and cognitive processes and so it is most directly concerned with Marr's second level. But any theory at the second level is *constrained from above* by the computational theory at the first level. An empirical proposal about how a computation is actually carried out will be incorrect if the proposed algorithm is not even a possible way, in principle, of accomplishing the transformation from input to output. Theories at the second level are also *constrained from below* by the theory of hardware implementation, or physical realization, at the third level. As Marr put it: 'Some types of algorithm will suit some physical substrates better than others' (Marr, 1982, p.24).

Marr's overall view, which has been influential in cognitive science and cognitive psychology, was that constraint from above, by the computational theory, took precedence over constraint from below, by the hardware implementation. Reflecting on neurophysiological investigations of vision in the 1970s, he became convinced that an important level of understanding had been missing:

> *There must exist an additional level of understanding at which the character of the information-processing tasks carried out during perception are analyzed and understood in a way that is independent of the particular mechanisms and structures that implement them in our heads. This was what was missing – the analysis of the problem as an information-processing task. Such analysis does not usurp an understanding at other levels – of neurons or of computer programs – but it is a necessary complement to them, since without it there can be no real understanding of the function of all those neurons.*
>
> (Marr, 1982, p.19)

The second reason why there has been relatively little discussion of the organization and function of the physical brain in cognitive psychology is that, until recently, there have been few techniques and tools that could be used to undertake the kinds of investigation at level 3 that would speak to issues concerning human cognitive functioning. In essence, experimental investigations were limited to animal studies, where invasive methods such as lesion studies and single-cell recording of neurons could take place. For ethical reasons, these kinds of studies could not be carried out on humans.

However, the situation has changed quite dramatically over recent years. Neuroimaging techniques such as positron emission tomography (PET), functional magnetic resonance imaging (fMRI), electroencephalography (EEG), and magnetoencephalography (MEG) allow metabolic, electrical, and magnetic correlates of neural activity to be measured, with varying degrees of spatial and temporal resolution, during the performance of cognitive tasks (Chapter 14). At the same time, there has been massive growth in connectionist modelling that is claimed to be 'brain-like' or at least neurally plausible. So the question of the relationship between the kinds of cognitive models discussed in this book and neural models that are pitched at Marr's level 3 has become a matter of significant theoretical interest and, indeed, dispute.

Contemporary theorists in the cognitive sciences commonly and reasonably assume that cognitive functioning is realized in neural functioning. So Marr is surely right that, in principle, things we find out about neural functioning might constrain models and theories of cognitive processes. This bottom-up constraint from level 3 to level 2 goes hand-in-hand with top-down guidance from level 2 to level 3. It is hard to see how an investigation at the level of the neural implementation of cognitive processes could ever get started without there first being a model (however sketchy) of the cognitive processes themselves.

If there is both bottom-up constraint and top-down guidance between Marr's second and third levels then we should expect reciprocal interaction between cognitive psychology and neurobiology, with challenges and insights flowing in both directions. This, in outline, is what Patricia Churchland (1986) describes as co-evolution of theories: 'theories at distinct levels often co-evolve … as each informs and corrects the other' (P.S. Churchland, 1986, p.284).

We should not, however, expect that drawing inferences about cognitive psychological theories from neurobiological findings will be easy or straightforward. A simple example illustrating this point is presented in Box 20.4.

5.2 Co-evolution and reduction

Patricia Churchland links co-evolution of theories with reduction of one theory to the other: 'the discoveries and problems of each theory may suggest modifications, developments, and experiments for the other, and thus the two evolve towards a reductive consummation' (P.S. Churchland, 1986, p.284). Even if it is not really obligatory that interdisciplinary interaction should aim at 'reductive consummation', the idea that theories in cognitive psychology *might* reduce to theories in neurobiology requires some discussion.

The topic of inter-theoretic reduction is a complex one (Churchland and Churchland, 1998 give a basic introduction). But the basic thought is that 'because everything in the world is made of the same basic stuff in complex combinations, the laws of biology ought to be derivable from those of chemistry, and the laws of chemistry from the laws of physics' (Ladyman, 2002, p.95). Similarly, the idea is that the laws of cognitive psychology ought to be logically derivable from laws that figure in neurobiological theories. Examples from the history of science that are usually adduced to illustrate and motivate this view are the reduction of genetics to molecular biology and the reduction of chemistry to quantum mechanics.

5.2.1 Reduction and the CMM

There is, however, a line of thought leading to the opposite claim, that cognitive psychology is not even a candidate for reduction to neurobiology. The CMM

BOX 20.4 Descending pathways and informational encapsulation

As we saw earlier in this chapter, Jerry Fodor argues that perception is informationally encapsulated from our beliefs and knowledge about the world. In one of a number of responses to this claim, the neurophilosopher Paul Churchland pointed out that, in addition to neuronal pathways that 'ascend' from the retina to primary visual cortex (via structures such as the lateral geniculate nucleus), there is evidence that there are also matched 'descending pathways' that 'lead us stepwise back through the intermediate brain areas and all the way out to the earliest processing systems at the retina' (P.M. Churchland, 1989, p.266). Churchland interprets these 'descending pathways' as feedback pathways from the central systems to the input systems that 'strongly suggest' that perceptual processes are *not* encapsulated.

Fodor's reaction to this neurobiological objection to informational encapsulation is instructive. He says: 'Heaven knows what psychological function "descending pathways" subserve. . . . One thing is clear: if there is no cognitive penetration of perception then at least "descending pathways" aren't for that' (Fodor, 1990, p.261). One way to interpret this remark is to read Fodor as denying, in principle, the relevance of neurobiological findings to cognitive psychology. But this would not be a credible interpretation, given Fodor's repeated insistence that there is no *a priori* way to put a boundary around the kinds of evidence that may be relevant to a theory. (Recall the isotropy of scientific confirmation discussed in Section 3.2.)

What Fodor is suggesting – in accord with the co-evolution view – is that *as matters currently stand* our understanding of the role of the descending pathways is too primitive to be taken as evidence against informational encapsulation. Moreover, we should expect that empirical theories in cognitive psychology (and, indeed, abstract computational theories) will guide our understanding of the role of those descending pathways. If we were to find good reasons for thinking that those pathways do carry information from central systems to input systems, then we might need to rethink our understanding of modules and encapsulation. But until then, assumptions of modularity and encapsulation may guide our understanding of the functions of descending pathways.

models 'the mind as the software of the brain' (Block, 1995). The same software, with the same scheme of representations and the same algorithmic processes (Marr's level 2), could run on different hardware (Marr's level 3). So the same cognitive psychological theory could be true of minds with very different physical realizations, some of them biological but some of them not. This point about the variable physical realization of cognitive states and processes blocks any straightforward reduction of cognitive psychology to neurobiology. A truly general cognitive psychological theory would apply to the cognitive states and processes of beings with wholly different physical compositions, some of which would not be biological at all. No neurobiological theory could match the generality of a psychological theory of that kind. Thus, there is a sense in which, as Block puts it, 'the computer model of the mind is profoundly *unbiological*' (1995, p.390).

This is an important line of thought but we should also notice its limitations. Variable physical realization blocks the reduction of completely general psychological theories, such as the CMM, to neurobiology. But this line of thought, by itself, leaves it open that a psychological theory of specifically *human* information processing might, in principle, be reduced to human neurobiology.

5.2.2 Reduction and the history of science

The linguist Noam Chomsky (2000, 2002) has articulated another problem for reductionism. He argues that inter-theoretic reduction is historically rare, and that where there have been genuine cases of reduction – such as that of chemistry to physics – this was only possible when there was a radical change in the more basic science – in this case, physics. The moral Chomsky draws is that reduction is neither to be aimed for nor to be expected.

The only sensible aim is for each theoretical enterprise to pursue its own path (as did chemistry and physics). It cannot be ruled out that there will be what Chomsky calls the 'unification' of theories (as with chemistry and physics), but nor can we be at all certain that human intelligence will be up to the task of unification. So far as psychology and neurobiology are concerned, 'one can entertain the idea that "the mental is the neurophysiological at a higher level", but for the present, only as a guide to enquiry, without much confidence about what "the neurophysiological" will prove to be' (Chomsky, 2003, p.265).

Chomsky's point can, perhaps, be put in the following way. The reduction of one theory to another requires that both theories are successful in their respective domains. If either (or both) is not antecedently successful, then why would one either expect or want reduction? With regard to psychology and neurobiology, our theories are tentative and far from being in the state where a reduction is, even in principle, in the offing. So, as things currently stand, the claim that psychology will reduce to neurobiology is merely a *historical speculation*, and one based on very scant evidence from the history of science.

5.3 Is cognitive psychology just a historical staging post?

Despite these reservations about reduction, it may still be tempting to think that contemporary cognitive psychology is a mere historical staging post *en route* to the terminus of a neurobiological theory of the mind. You might expect that cognitive psychology will disappear when that end point is reached. Philosophers Ian Gold and Daniel Stoljar have described this view as a version of *the neuron doctrine*: 'a successful theory of the mind will be a wholly neuroscientific theory' (Gold and Stoljar, 1999, p.809).

The neuron doctrine has different versions depending on how the term 'neuroscientific' is interpreted. If a neuroscientific theory is a theory in *cognitive* neuroscience, drawing on some combination of concepts from cognitive psychology and neurobiology (or, more generally, from the psychological and biological sciences), then the neuron doctrine is somewhat bland and uncontroversial. But if a neuroscientific theory is a theory that draws only on neuro*biology* then we have a much more radical neuron doctrine: 'a successful theory of the mind will be a theory of the brain expressed *in terms of* the basic structural and functional properties of neurons, ensembles or structures' (Gold and Stoljar, 1999, p.814). In a little more detail, the radical neuron doctrine says:

> *neurophysiology, neuroanatomy, and neurochemistry will by themselves eventually have the conceptual resources to understand the mind and, as a consequence, a successful theory of the mind will make no reference to anything like the concepts of . . . the psychological sciences as we currently understand them.*
>
> (Gold and Stoljar, 1999, p.814)

On this view, we shall have not the co-evolution of theories, but rather the gradual extinction of cognitive psychological theories as neurobiological theories take over.

A temptation to believe the radical neuron doctrine may arise from an assumption that (as we noted earlier) most cognitive psychologists and cognitive scientists share; namely, that cognitive functioning is realized in neural functioning. It may seem obvious that, if mental functioning is the result of the physical organization and functioning of the brain, then a theory of the mind will ultimately be produced that talks only about the brain – about neurons, neural circuits, neurotransmitters, and so on.

But this would be to move much too fast. To use an example of Gold and Stoljar's (1999, p.815): earthquakes are made up of the movement of physical particles that behave in accordance with the laws of physics. But the science of earthquakes shows no signs of being replaced by physics, and the reason for this is quite general. Just because phenomena of some kind, Xs, are made up of lots of Ys, it does not follow that we will achieve an understanding of Xs in terms of the science of Ys. We should not, then, allow ourselves to slide from the widely held assumption that the mind is realized in the brain to the radical and controversial version of the neuron doctrine. Furthermore, if we were to begin on this slide then it would be completely unmotivated to maintain that the *real* science of the mind is neurobiology. For neural phenomena are made up of the activity of lots of molecules and atoms, protons, neutrons, and electrons, and ultimately, of whatever turn out to be the fundamental building blocks of all matter.

5.4 Is cognitive psychology autonomous from neurobiology?

Philosophical commentators on the relationship between cognitive psychology and neurobiology have used two different notions of autonomy, each of them important.

One notion of autonomy is *irreducibility*. Suppose, as we assume is true, and this book is predicated on, that theories in cognitive psychology provide genuine empirical explanations. If the laws that figure in those theories are not derivable from laws in neurobiology then cognitive psychology is autonomous from neurobiology in this first sense (Fodor, 1998, p.9). If, instead, the key concepts in cognitive psychology can be defined in terms of concepts in neurobiology, and if the laws that figure in cognitive psychological theories are derivable from laws of neurobiology, then cognitive psychology is reducible and not autonomous in that same sense.

Suppose, for a moment, that cognitive psychology were to be reducible to neurobiology. It is important to notice that it would not immediately follow that the radical neuron doctrine was also true. According to the radical neuron doctrine, cognitive psychology will, in due course, be discarded. But a successful reduction can be a way of vindicating, rather than eliminating, the reduced theory. The Churchlands, who are certainly friends of reduction, say: 'The discipline of psychology will still be with us a hundred years from now, and five hundred, and a thousand. . . . Its existence is unthreatened' (P.M. Churchland and P.S. Churchland, 1996, p.220).

A second notion of autonomy is *independence*. Cognitive psychology is autonomous from neurobiology in this second sense if it is entirely unconstrained by neurobiology. Note that autonomy in this second sense entails autonomy in the first: if cognitive psychology were to be fully independent from neurobiology then it could not be reducible to neurobiology. However, the reverse does not hold: cognitive psychology could be constrained by neurobiology – as in Marr's account – without being reducible to neurobiology.

We can therefore reject the autonomy of cognitive psychology in the second sense without rejecting autonomy in the first sense. We can allow that theories in cognitive psychology and neurobiology co-evolve without being committed to reductive consummation. Indeed, this seems to be quite an attractive account of the relationship between the two disciplines (Stone and Davies, 1999).

Some psychologists maintain that the cognitive psychological level (Marr's level 2) takes *priority* over the neurobiological level (Marr's level 3). This priority has theoretical and practical aspects. The theoretical aspect of the priority of cognitive psychology is expressed by Coltheart and Langdon, when they say that 'it can be very hard to understand what a system is actually doing if one's only information about it is a description at the physical-instantiation level' (1998, p.150). This aspect of the priority of cognitive psychology seems to coincide with that we earlier described as top-down guidance from Marr's level 2 to level 3. In Marr's account, the project at level 3 is to discover how the representations and algorithms *specified at level 2* are physically realized in the brain. It would seem to be a very much more difficult project to work out which representations and algorithms were being realized, given only a neurobiological description of neural states and processes.

The practical aspect of the priority of cognitive psychology is simply a manifestation of the fact that all scientists need to make decisions about how to conduct research with limited resources. One option for a cognitive psychologist may be to push ahead with the development of an information-processing model while abstracting, for the time being, from debates within neurobiology. For example, Andy Young describes how the development of a cognitive model of face recognition proceeded in relative abstraction from debate about a neuroanatomical proposal (Bauer, 1984) concerning two of the routes in the model. Young comments: 'It is a good example of how, even though they can often be usefully combined, psychological and neurological hypotheses can have some degree of independence from each other' (Young, 1998, p.44). Cognitive psychologists who make these practical decisions would typically acknowledge, of course, that the model they develop in abstraction from neurobiology is still potentially constrained from below. The model would have to be rejected if there were no neurobiological account consistent with it.

Sometimes, however, these claims about the theoretical and practical priority of cognitive psychology seem to be associated with the less plausible claim that neurobiological findings are irrelevant to cognitive models. Examples of findings whose relevance sometimes seems to be contested include findings about the location of neural activity in neurologically healthy individuals who are performing a cognitive task, or the location of lesions in patients who can no longer perform a particular task. As was suggested in Chapter 1, the relevance of neuroimaging to cognitive psychology has thus become a topic of lively discussion and debate (Henson, 2005; Coltheart, 2006b).

SUMMARY OF SECTION 5

- According to Marr's account of levels of explanation, models in cognitive psychology are constrained from above by the computational theory, and from below by the theory of hardware implementation. It was also part of the account that constraint from above takes precedence over constraint from below.
- On the co-evolution of theories view, there is reciprocal interaction between cognitive psychology and neurobiology, with challenges and insights flowing in both directions.
- There are theoretical and historical grounds to doubt whether cognitive psychology will be reduced to neurobiology.
- The radical neuron doctrine predicts that cognitive psychology will eventually be replaced by neurobiology, but this doctrine is unmotivated.
- Claims about the theoretical and practical priority of cognitive psychology are sometimes combined with less plausible claims about the irrelevance of neurobiological findings to cognitive models.

6 CONCLUSION

This chapter has introduced you to some of the main theoretical debates in cognitive psychology and in cognitive science more broadly. The first debate (Section 2) was about rules and representations and the connectionist challenge to the computational model of the mind (CMM). One strand in this large-scale debate concerned the competing connectionist and rule-based accounts of a particular cognitive process, the formation of the past tense of verbs. The second large-scale debate (Section 3) was about the modularity of the mind and about challenges to the two parts – positive and negative – of Fodor's account of modularity. One important question was whether language processing, and specifically the process of lexical access, is subject to top-down influence from non-linguistic knowledge.

In Section 4, we introduced the debate about the role of double dissociation evidence in cognitive neuropsychology. In this context, returning to the past tense and to connectionist models of past-tense formation allowed us to highlight two points. First, cognitive neuropsychology, like all of normal science, proceeds by inference to the best explanation. Second, competing models of cognitive processes should be assessed not only for how well they fit a particular set of data, but also for how well they explain the available evidence. In the final section, we briefly introduced debates about the relationship between cognitive psychology and neurobiology – reduction, co-evolution, priority, and independence. The contribution that neuroimaging can make to cognitive psychology is a topic of continuing and widespread discussion.

None of these debates has a resolution that carries a consensus amongst cognitive psychologists and cognitive scientists. But perhaps there would be widespread agreement that the development of explicit, implemented computational models – including connectionist models – has made a substantial contribution to cognitive psychology over the last 25 years. And perhaps the success of computational cognitive psychology serves to highlight the question of whether, as Fodor claims, some of the aspects of human cognition that are the most important to us are computationally intractable.

There are, of course, many other theoretical debates that warrant attention. Among those that we have most reluctantly omitted are debates about the cognitive basis of our everyday psychological understanding of ourselves and others (theory of mind; Davies and Stone, 1995), about the prospects for a science of consciousness (Chapter 18; Davies and Humphreys, 1993; Weiskrantz and Davies, 2008), and about the extension of the methods of cognitive neuropsychology to disorders, such as delusions, that were previously regarded as psychiatric phenomena (Chapter 15; Stone and Young, 1997).

We hope that you will take away from this chapter a greater insight into the fundamental principles that lie behind theories in contemporary cognitive psychology, and that this will enhance your understanding of the individual topics covered in this book.

Dedication

Tony Stone, the author of the original version of this chapter (published in the first edition of this book in 2005), died in June 2010 at the age of 52. This revised chapter is dedicated to his memory.

FURTHER READING

When we introduced the computational model of the mind, we adopted Ned Block's phrase, 'the mind as the software of the brain'. Block's chapter of that title provides an excellent account of the CMM, as does Tim Crane's fine book. Andy Clark also provides an extensive introduction to computation and connectionism. The collection edited by Cohen, Johnston, and Plunkett is a valuable resource for thinking about the way in which computational cognitive psychology, and particularly connectionist modelling, impacts on cognitive neuropsychology. The relevance (or not) of neuroimaging to cognitive psychology is discussed in a Forum in the journal Cortex (2006), with the target paper by Max Coltheart.

Block, N. (1995) 'The mind as the software of the brain', in Smith, E.E. and Osherson, D.N. (eds.) *An Invitation to Cognitive Science*, *Volume 3*: *Thinking*, Cambridge, MA, MIT Press. Ned Block's classic paper explains the computational model of the mind in detail, and includes informative discussions of machine intelligence, variable physical realization, the language of thought, explanatory levels, and reduction.

Clark, A. (2001) *Mindware: An Introduction to the Philosophy of Cognitive Science*, Oxford, Oxford University Press. Andy Clark's engaging and exciting book begins with 'mindware as software', continues with classical symbol manipulation and connectionism, and introduces further topics, including robots, dynamical systems, and embodied and embedded cognition.

Cohen, G., Johnston, R.A., and Plunkett, K. (eds.) (2000) *Exploring Cognition: Damaged Brains and Neural Networks. Readings in Cognitive Neuropsychology and Connectionist*

Modelling, Hove, Psychology Press. This collection, with an excellent introductory chapter by Robert Johnston and Nick Braisby, includes papers by Coltheart and colleagues (1993), Juola and Plunkett (2000), and Marslen-Wilson and Tyler (1998), along with a wealth of additional resources.

Coltheart, M. (2006b) 'What has functional neuroimaging told us about the mind (so far)?' *Cortex*, vol.42, pp.323–31. In his position paper, Max Coltheart argues that 'no functional neuroimaging research to date has yielded data that can be used to distinguish between competing psychological theories'; this is followed by comments and a response.

Crane, T. (2003) *The Mechanical Mind: A Philosophical Introduction to Minds, Machines and Mental Representations*, 2nd edition, London, Routledge. Tim Crane's excellent book explains representation, the computational model of the mind, and connectionism; it includes brief discussions of systematicity and modularity and concludes with a chapter on consciousness.

REFERENCES

Barkow, J.H., Cosmides, L., and Tooby, J. (eds.) (1992) *The Adapted Mind: Evolutionary Psychology and the Generation of Culture*, Oxford, Oxford University Press.

Baron-Cohen, S., Tager-Flusberg, H., and Cohen, D. (eds.) (1999) *Understanding Other Minds: Perspectives from Developmental Cognitive Neuroscience*, 2nd edition, Oxford, Oxford University Press.

Bauer, R.M. (1984) 'Autonomic recognition of names and faces in prosopagnosia: a neuropsychological application of the guilty knowledge test', *Neuropsychologia*, vol.22, pp.457–69.

Block, N. (1995) 'The mind as the software of the brain', in Smith, E.E. and Osherson, D.N. (eds.) *An Invitation to Cognitive Science: Volume 3 – Thinking*, Cambridge, MA, MIT Press.

Bruce, V. and Young, A. (1986) 'Understanding face recognition', *British Journal of Psychology*, vol.77, pp.305–27.

Caramazza, A. (1986) 'On drawing inferences about the structure of normal cognitive systems from the analysis of patterns of impaired performance: the case for single-patient studies', *Brain and Cognition*, vol.5, pp.41–66.

Carruthers, P. (2006) *The Architecture of the Mind: Massive Modularity and the Flexibility of Thought*, Oxford, Oxford University Press.

Chomsky, N. (2000) *New Horizons in the Study of Language and Mind*, Cambridge, Cambridge University Press.

Chomsky, N. (2002) *On Nature and Language*, Cambridge, Cambridge University Press.

Chomsky, N. (2003) 'Replies', in Antony, L.M. and Hornstein, N. (eds.) *Chomsky and his Critics*, Oxford, Blackwell.

Churchland, P.M. (1989) *A Neurocomputational Perspective: The Nature of Mind and the Structure of Science*, Cambridge, MA, MIT Press.

Churchland, P.M. and Churchland, P.S. (1996) 'Replies from the Churchlands', in McCauley, R.N. (ed.) *The Churchlands and their Critics*, Oxford, Blackwell Publishers.

Churchland, P.M. and Churchland, P.S. (1998) 'Intertheoretic reduction: a neuroscientist's field guide', in Churchland, P.M. and Churchland, P.S. *On the Contrary: Critical Essays*, 1987–1997, Cambridge, MA, MIT Press.

Churchland, P.S. (1986) *Neurophilosophy: Toward a Unified Science of the Mind/Brain*, Cambridge, MA, MIT Press.

Coltheart, M. (1999) 'Modularity and cognition', *Trends in Cognitive Sciences*, vol.3, pp.115–20.

Coltheart, M. (2001) 'Assumptions and methods in cognitive neuropsychology', in Rapp, B. (ed.) *The Handbook of Cognitive Neuropsychology: What Deficits Reveal About the Human Mind*, Hove, Psychology Press.

Coltheart, M. (2006a) 'Acquired dyslexias and the computational modelling of reading', in Coltheart, M. and Caramazza, A. (eds.) *Cognitive Neuropsychology Twenty Years On*, Hove, Psychology Press.

Coltheart, M. (2006b) 'What has functional neuroimaging told us about the mind (so far)?' *Cortex*, vol.42, pp.323–31.

Coltheart, M. and Davies, M. (2003) 'Inference and explanation in cognitive neuropsychology' *Cortex*, vol.39, pp.188–91.

Coltheart, M. and Langdon, R. (1998) 'Autism, modularity and levels of explanation in cognitive science', *Mind & Language*, vol.13, pp.138–52.

Coltheart, M., Curtis, B., Atkins, P., and Haller, M. (1993) 'Models of reading aloud: dual-route and parallel-distributed-processing approaches', *Psychological Review*, vol.100, pp.589–608. (Reprinted in Cohen, G., Johnston, R.A., and Plunkett, K. (eds.) *Exploring Cognition: Damaged Brains and Neural Networks. Readings in Cognitive Neuropsychology and Connectionist Modelling*, Hove, Psychology Press, 2000.)

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

Cosmides, L. and Tooby, J. (1994) 'Origins of domain specificity: the evolution of functional organization', in Hirschfeld, L.A. and Gelman, S.A. (eds.) *Mapping the Mind: Domain

Specificity in Cognition and Culture, Cambridge, Cambridge University Press.

Davies, M. (1995) 'Two notions of implicit rules', in Tomberlin, J. (ed.) *Philosophical Perspectives, Volume 9: AI, Connectionism and Philosophical Psychology*, Oxford, Blackwell.

Davies, M. (2010) 'Double dissociation: understanding its role in cognitive neuropsychology', *Mind & Language*, vol.25, pp.500–40.

Davies, M. and Humphreys, G.W. (eds.) (1993) *Consciousness: Psychological and Philosophical Essays*, Oxford, Blackwell Publishers.

Davies, M. and Stone, T. (eds.) (1995) *Folk Psychology: The Theory of Mind Debate*, Oxford, Blackwell Publishers.

Dennett, D.C. (1984) 'Cognitive wheels: the frame problem of AI', in Hookway, C. (ed.) *Minds, Machines, and Evolution: Philosophical Studies*, Cambridge, Cambridge University Press. (Reprinted in Boden, M.A. (ed.) *The Philosophy of Artificial Intelligence*, Oxford, Oxford University Press, 1990.)

Ellis, A.W. and Young, A.W. (1998) *Cognitive Neuropsychology: A Textbook with Readings*, Hove, Psychology Press.

Fodor, J.A. (1975) *The Language of Thought*, Cambridge, MA, Harvard University Press.

Fodor, J.A. (1981) 'Some notes on what linguistics is about', in Block, N. (ed.) *Readings in the Philosophy of Psychology, Volume II*, Cambridge, MA, Harvard University Press. (Reprinted in Katz, J. (ed.) *The Philosophy of Linguistics*, Oxford, Oxford University Press, 1985.)

Fodor, J.A. (1983) *The Modularity of Mind*, Cambridge, MA, MIT Press.

Fodor, J.A. (1985) 'Précis of *The Modularity of Mind*', *Behavioral and Brain Sciences*, vol.8, no.1, pp.1–42. (Reprinted in Fodor, 1990, pp.195–206.)

Fodor, J.A. (1987) *Psychosemantics: The Problem of Meaning in the Philosophy of Mind*, Cambridge, MA, MIT Press.

Fodor, J.A. (1990) *A Theory of Content and Other Essays*, Cambridge, MA, MIT Press.

Fodor, J.A. (1998) *In Critical Condition*, Cambridge, MA, MIT Press.

Fodor, J.A. (2000) *The Mind Doesn't Work that Way: The Scope and Limits of Computational Psychology*, Cambridge, MA, MIT Press.

Fodor, J.A. (2005) 'Reply to Steven Pinker "So how *does* the mind work?"' *Mind & Language*, vol.20, pp.25–32.

Fodor, J.A. (2008) *LOT 2: The Language of Thought Revisited*, Oxford, Oxford University Press.

Gallistel, C.R. (2001) 'Mental representations, psychology of', in Smelser, N.J. and Baltes, P.B. (eds.) *International Encyclopedia of the Social and Behavioral Sciences*, Amsterdam and New York, Elsevier.

Garson, J. (2010) 'Connectionism', in Zalta, E.N. (ed.) *The Stanford Encyclopedia of Philosophy* [online] Available from: http://plato.stanford.edu/entries/connectionism/

Gold, I. and Stoljar, D. (1999) 'A neuron doctrine in the philosophy of neuroscience', *Behavioral and Brain Sciences*, vol.22, pp.809–69.

Henson, R.N.A. (2005) 'What can functional neuroimaging tell the experimental psychologist?' *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, vol.58, pp.193–233.

Joanisse, M.F. and Seidenberg, M.S. (1999) 'Impairments in verb morphology after brain injury: a connectionist model', *Proceedings of the National Academy of Science*, vol.96, pp.7592–7.

Joanisse, M.F. and Seidenberg, M.S. (2005) 'Imaging the past: neural activation in frontal and temporal regions during regular and irregular past-tense processing', *Cognitive, Affective and Behavioral Neuroscience*, vol.5, pp.282–96.

Juola, P. and Plunkett, K. (2000) 'Why double dissociations don't mean much', in Cohen, G., Johnston, R.A., and Plunkett, K. (eds.) *Exploring Cognition: Damaged Brains and Neural Networks. Readings in Cognitive Neuropsychology and Connectionist Modelling*, Hove, Psychology Press.

Ladyman, J. (2002) *Understanding Philosophy of Science*, London, Routledge.

Lashley, K.S. (1930) 'Basic neural mechanisms in behavior', *Psychological Review*, vol.37, pp.1–24.

Marr, D. (1982) *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*, New York, W.H. Freeman and Company. (Reprinted Cambridge, MA, MIT Press, 2010). (Chapter 1, 'The philosophy and the approach', reprinted in Yantis, S. (ed.) *Visual Perception: Essential Readings*, Hove, Psychology Press, 2003.)

Marslen-Wilson, W. and Tyler, L.K. (1987) 'Against modularity', in Garfield, J. (ed.) *Modularity in Knowledge Representation and Natural Language Understanding*, Cambridge, MA, MIT Press.

Marslen-Wilson, W. and Tyler, L.K. (1997) 'Dissociating types of mental computation', *Nature*, vol.387, pp.592–4.

Marslen-Wilson, W. and Tyler, L.K. (1998) 'Rules, representations, and the English past tense', *Trends in Cognitive Sciences*, vol.2, pp.428–35. (Reprinted in Cohen, G., Johnston, R.A., and Plunkett, K. (eds.) *Exploring Cognition: Damaged Brains and Neural Networks. Readings in Cognitive Neuropsychology and Connectionist Modelling*, Hove, Psychology Press, 2000.)

Marslen-Wilson, W., Brown, C.M., and Tyler, L.K. (1988) 'Lexical representations in spoken language comprehension', *Language and Cognitive Processes*, vol.3, pp.1–16.

Marslen-Wilson, W., Tyler, L.K., and Koster, C. (1993) 'Integrative processes in utterance resolution', *Journal of Memory and Language*, vol.32, pp.647–66.

McClelland, J.L. and Patterson, K. (2002a) '"Words or rules" cannot exploit the regularity of exceptions', *Trends in Cognitive Sciences*, vol.6, pp.464–5.

McClelland, J.L. and Patterson, K. (2002b) 'Rules or connections in past-tense inflections: what does the evidence rule out?' *Trends in Cognitive Sciences*, vol.6, pp.465–72.

McClelland, J.L., Rumelhart, D.E., and Hinton, G.E. (1986) 'The appeal of parallel distributed processing', in Rumelhart, D.E. and McClelland, J.L. (eds.) *Parallel Distributed Processing: Explorations in the Microstructure of Cognition, Volume 1: Foundations*, Cambridge, MA, MIT Press.

Pinker, S. (1997) *How the Mind Works*, New York, W.W. Norton.

Pinker, S. (2005a) 'So how *does* the mind work?' *Mind & Language*, vol.20, pp.1–24.

Pinker, S. (2005b) 'A reply to Jerry Fodor on how the mind works', *Mind & Language*, vol.20, pp.33–8.

Pinker, S. (2006) 'Whatever happened to the past tense debate?' in Baković, E., Ito, J., and McCarthy, J.J. (eds.) *Wondering at the Natural Fecundity of Things: Essays in Honor of Alan Prince*, Santa Cruz, CA, Linguistics Research Center, University of California, Santa Cruz. Available online from:

http://www.escholarship.org/uc/item/0xf9q0n8

Pinker, S. and Prince, A. (1988) 'On language and connectionism; analysis of a parallel distributed processing model of language acquisition', *Cognition*, vol.28, pp.73–193.

Pinker, S. and Ullman, M.T. (2002a) 'The past and future of the past tense', *Trends in Cognitive Sciences*, vol.6, pp.456–63.

Pinker, S. and Ullman, M.T. (2002b) 'Combination and structure, not gradedness, is the issue', *Trends in Cognitive Sciences*, vol.6, pp.463–74.

Plaut, D.C., McClelland, J.L., Seidenberg, M.S., and Patterson, K. (1996) 'Understanding normal and impaired word reading: computational principles in quasi-regular domains', *Psychological Review*, vol.103, pp.56–115.

Plunkett, K. and Bandelow, S. (2006) 'Stochastic approaches to understanding dissociations in inflectional morphology', *Brain and Language*, vol.98, pp.194–209.

Plunkett, K. and Marchman, V. (1993) 'From rote learning to system building: acquiring verb morphology in children and connectionist nets', *Cognition*, vol.48, pp.21–69.

Robbins, P. (2009) 'Modularity of mind', in Zalta, E.N. (ed.) *The Stanford Encyclopedia of Philosophy* [online] Available from: http://plato.stanford.edu/entries/modularity-mind/

Rumelhart, D.E. and McClelland, J.L. (1986) 'On learning the past tenses of English verbs', in McClelland, J.A. and Rumelhart, D.E. (eds.) *Parallel Distributed Processing: Explorations in the Microstructure of Cognition, Volume 2: Psychological and Biological Models,* Cambridge, MA, MIT Press.

Searle, J.R. (1992) *The Rediscovery of the Mind*, Cambridge, MA, MIT Press.

Seidenberg, M.S. and McClelland, J.L. (1989) 'A distributed, developmental model of word recognition and naming', *Psychological Review*, vol.96, pp.523–68.

Seidenberg, M.S. and Plaut, D.C. (2006) 'Progress in understanding word reading: data fitting versus theory building', in Andrews, S. (ed.) *From Inkmarks to Ideas: Current Issues in Lexical Processing*, Hove, Psychology Press.

Shallice, T. (1988) *From Neuropsychology to Mental Structure*, Cambridge, Cambridge University Press.

Shanahan, M. (2009) 'The frame problem', in Zalta, E.N. (ed.) *The Stanford Encyclopedia of Philosophy* [online] Available from: http://plato.stanford.edu/entries/frame-problem/

Sperber, D. (1994) 'The modularity of thought and the epidemiology of representations', in Hirschfeld, L.A. and Gelman, S.A. (eds.) *Mapping the Mind: Domain Specificity in Cognition and Culture*, Cambridge, Cambridge University Press.

Sperber, D. (1996) *Explaining Culture: A Naturalistic Approach*, Oxford, Blackwell.

Sperber, D. (2002) 'In defense of massive modularity', in Dupoux, I. (ed.) *Language, Brain and Cognitive Development: Essays in Honor of Jacques Mehler*, Cambridge, MA, MIT Press.

Stone, T. and Davies, M. (1999) 'Autonomous psychology and the moderate neuron doctrine', *Behavioral and Brain Sciences*, vol.22, pp.849–50.

Stone, T. and Young, A.W. (1997) 'Delusions and brain injury: the philosophy and psychology of belief', *Mind & Language*, vol.12, pp.327–64.

Swinney, D.A. (1979) 'Lexical access during sentence comprehension: (re)consideration of context effects', *Journal of Verbal Learning and Verbal Behavior*, vol.18, no.5, pp.645–59.

Tyler, L.K. and Marslen-Wilson, W. (1977) 'The on-line effects of semantic context on syntactic processing', *Journal of Verbal Learning and Verbal Behavior*, vol.16, pp.683–92.

Tyler, L.K. and Marslen-Wilson, W. (1982) 'Processing utterances in discourse contexts: on-line resolution of anaphors', *Journal of Semantics*, vol.1, pp.297–314.

Ullman, M.T. (2001) 'A neurocognitive perspective on language: the declarative/procedural model', *Nature Reviews Neuroscience*, vol.2, pp.717–26.

Ullman, M.T., Corkin, S., Coppola, M., Hickol, G., Growdon, J.H., Koroshetz, W.J., and Pinker, S. (1997) 'A neural dissociation within language: evidence that the mental dictionary is part of declarative memory, and that grammatical rules are processed by the procedural system', *Journal of Cognitive Neuroscience*, vol.9, pp.266–76.

Ullman, M.T., Pancheva, R., Love, T., Lee, E., Swinney, D., and Hickok, G. (2005) 'Neural correlates of lexicon and grammar: evidence from the production, reading, and judgment of inflection in aphasia', *Brain and Language*, vol.93, pp.185–238.

Weiskrantz, L. and Davies, M. (eds.) (2008) *Frontiers of Consciousness: Chichele Lectures*, Oxford, Oxford University Press.

Woollams, A.M., Joanisse, M.F., and Patterson, K. (2009) 'Past-tense generation from form versus meaning: behavioural data and simulation evidence', *Journal of Memory and Language*, vol.61, pp.55–76.

Young, A.W. (1998) *Face and Mind*, Oxford, Oxford University Press.

Visit www.oxfordtextbooks.co.uk/orc/braisby2e/ for online resources accompanying this chapter.