EDITORIAL

WHAT HAS FUNCTIONAL NEUROIMAGING TOLD US ABOUT THE MIND (SO FAR)?

(Position paper presented to the European Cognitive Neuropsychology Workshop, Bressanone, 2005)

Max Coltheart

(Macquarie Centre for Cognitive Science, Macquarie University, Sydney, Australia)

One way in which scientists currently study cognition is by cognitive neuroimaging – by recording neural activity in a person's brain while the recorded person is carrying out some cognitive activity. There are numerous different reasons for doing this kind of work. I will consider only one of these reasons, namely, to try to learn more about cognition itself. All other motivations for doing cognitive neuroimaging (e.g., to seek to localise specific cognitive functions in specific brain regions) are beyond the scope of this paper.

Although there exists a huge volume of recent literature reporting the results of cognitive neuroimaging studies, there are surprisingly few papers which have evaluated this technique as a way of studying cognition itself. Some of these papers offer rather negative conclusions. Some of these negative conclusions are modest in scope ('What has functional neuroimaging told us about the mind so far?', 'Nothing so far in one cognitive domain, viz., language'; Poeppel, 1996), whereas others are much more sweeping ('What has functional neuroimaging told us about the mind so far?', 'Nothing, and it never will: the nature of cognition is such that this technique in principle cannot provide evidence about the nature of cognition'; van Orden and Paap, 1997; see also Uttal, 2001).

Henson (2005) has provided an invaluable framework for considering the role of brain-imaging data in cognitive psychology. He writes: "My main argument is that, provided one makes the assumption that there is some 'systematic' mapping from psychological function to brain structure, then functional neuroimaging data simply comprise another dependent variable, along with behavioural data, that can be used to distinguish between competing psychological theories" (p. 194).

I want to challenge this argument directly. I fully accept Henson's assumption that there is some systematic mapping from psychological function to brain structure. Nevertheless, I'll claim that no functional neuroimaging research to date has yielded data that can be used to distinguish between competing psychological theories. I emphasize that the scope of my claim is much narrower than the scope of Henson's. He is considering whether functional neuroimaging data *can ever* (i.e., in principle) be used to distinguish between competing psychological theories. I will be considering only whether functional neuroimaging data *has already* been successfully used to distinguish between competing psychological theories. Are there already clear examples of such successful use? If there are not, that of course does not imply that there won't be in the future. However, given the enormous volume of published recent empirical work in this area, if it turns out that none of this work can be used to distinguish between competing psychological theories, the in-principle question of whether cognitive neuroimaging data can ever serve this function will deserve much more attention than it has so far been given. But I will not be discussing this in-principle question: this paper is solely concerned with an in-practice so-far question: What have we learned so far about the mind from cognitive-neuroimaging research?

Three Disclaimers

Firstly, as mentioned above I will not be considering one use to which functional neuroimaging is often put, namely, the localization of cognitive functions in specific brain regions. That is a different enterprise from seeking to distinguish between competing psychological theories, and a paper about whether functional neuroimaging has successfully been used to localize some cognitive functions in the brain would be a different paper from the one I seek to write. My paper, like Henson's, is concerned solely with the impact of functional neuroimaging on the evaluation of theories that are *expressed solely at the psychological level.*

Secondly, even if one accepted the view that no work has yet been done in which functional neuroimaging data has successfully been used to distinguish between competing psychological theories, that does not mean such work might not be done in the future. As I have said above, I will not be concerned with this in-principle question,

but solely with an in-practice question: has such work already been done? Do we already have examples of such work?

Thirdly, I will not seek to review the entire body of work on functional cognitive neuroimaging – clearly, an impossible (or at least a book-length) task. Instead, my strategy will be to make my claim as clear as possible to everyone, and then to invite my readers to provide examples which they consider show me to be wrong. This makes my task a possible one.

What does "Distinguish between Competing Psychological Theories" Mean?

Let T_a and T_b be competing psychological theories. They are competing in the sense that they offer different accounts of the same psychological phenomena, so that at most only one of the theories could be true. One can't hope for data to show which theory is certainly true and which is certainly false; science isn't like that. But one can hope for data that makes one theory seem a sounder bet than the other; even better, if there several different results that all make one of the theories seem the sounder bet, and there are no results favouring the other theory, it is rational to prefer the universally supported theory. That is all I will mean by data "distinguishing between competing psychological theories".

It seems clear that behavioural data can serve to distinguish between competing psychological theories in this sense. I will give two examples where I consider that this has demonstrably occurred.

Serial versus Parallel Processing in Reading

T_a: according to this theory, all of the processes used to convert print to speech operate in parallel on the letters of the input string – there is no serial processing involved in reading aloud (Seidenberg and McClelland, 1989; Plaut et al., 1996; Harm and Seidenberg, 1999, 2004).

T_b: according to this theory, at least some of the processes used to convert print to speech operate serially left-to-right on the letters of the input string (Coltheart, 1978; Coltheart et al., 1993, 2001).

Some predictions which follow from T_b are:

(a) When a word contains an irregular grapheme-phoneme correspondence, the later in that word that correspondence is the less the word's reading-aloud latency will be affected by its irregularity.

(b) When the task is naming the colour in which nonwords are written, colour naming latency will be shorter when the first phoneme of the non word is the first phoneme of the colour name than when the two phonemes are different, and this facilitation effect will be smaller when the last phoneme of the nonword is same *versus* different.

(c) In a masked priming experiment where brief prime word and following target word share just one phoneme, the priming of target word RT will be larger when the shared phoneme is the first phoneme of the prime than when it is the last.

T_a predicts that none of these effects will occur; T_b predicts that all will. All three effects have been reported in the literature (see Rastle and Coltheart, in press, for a general review of this literature). And there are no reported data which conform to predictions from T_a and conflict with predictions from T_b. This set of results strongly favours T_b over T_a.

Semantic Contributions to Reading Aloud

T_a: according to this theory, irregular words, especially if they are low in frequency, require the assistance of access to semantics to be read aloud correctly. Plaut et al. (1996) and Rogers et al. (2004) have proposed T_a.

T_b: according to this theory, irregular-word reading can be accomplished correctly without any contribution from access to semantics. Goodall and Phillips (1995), Patterson and Shewell (1987), Lytton and Brust (1989), Coltheart et al. (2001) and others have proposed T_b

Patients with severe impairments of the semantic system or of access to it from print are relevant here. T_a predicts that *all* patients with such impairments will also be impaired at irregular word reading; T_b predicts that there will be patients with semantic impairments yet normal irregular-word reading. At least four different studies have documented patients with severe semantic impairment but normal accuracy in irregular-word reading (Blazely et al., 2005; Cipolotti and Warrington, 1995; Gerhand, 2001; Lambon-Ralph et al., 1995). This set of results strongly favours T_b over T_a.

These two examples are meant to illustrate the sense in which I consider behavioural data to have been successfully used to distinguish between psychological theories. My aim in this paper is to consider whether functional neuroimaging data have been successfully used to distinguish between psychological theories, in just the same sense.

The Approach

As I have said, I am concerned specifically here with distinguishing between psychological theories. Hence in discussing any functional neuroimaging work in any cognitive domain, I must always begin by stating two or more psychological theories T_a T_b … concerning that domain[1]. Then I can consider

[1]As Henson (2005) does when discussing the general form of inference from neuroimaging data to psychological theory.

whether there has been any neuroimaging work that has yielded data which provides good reason to favour one of these theories over the others. So what's needed is to show that T_a predicts X whilst T_b predicts ~X, where X is some pattern of neuroimaging data, and then to show that there exists functional neuroimaging work which demonstrates that X is the case or demonstrates that ~X is the case. When this happens, we have a situation in which functional neuroimaging data has successfully been used to distinguish between psychological theories. This is the approach Henson (2005) has taken in his paper; and he gives seven examples where he considers such success to have been achieved. I consider these examples in turn.

Example 1 – Recognition Memory: the Remember/Know Paradigm

Subjects are given some task (the Study Task) to carry out with a set of stimuli. After that task has been completed, their memory for the Study Task stimuli is tested via an Old/New recognition memory test; they have not been warned that this will happen. When subjects think a stimulus in the recognition memory test had been presented in the Study Task – i.e., when they respond "Old" in the Recognition Task – they are asked whether they think this because they recollect a particular aspect of the prior encounter ('Remember') or because the stimulus just seems familiar ('Know').

We seek a theory that can account for what makes subjects say "Remember" on some trials where they have correctly responded "Old" and what makes them say "Know" on other trials where they have correctly responded "Old".

Henson describes two such theories.

(1) T_a is dual-process theory (Yonelinas, 2002): "Remember" judgements entail recollection (and possibly familiarity judgement too) whereas "Know" judgements entail familiarity judgement but not recollection.

(2) T_b is the single-process model (Heathcote, 2003 and others): "Remember" and "Know" judgements simply reflect different response criteria along a single continuum of memory strength. No distinction need be made between one cognitive process called "recollection" and a second, distinct, cognitive process called "familiarity judgement". Instead, the subject just responds "Remember" when the memory trace is strong and "Know" when the memory trace is weak.

Henson et al. (1999) reported an fMRI study of subjects performing this Remember/Know recognition memory task. In the study condition, subjects performed a lexical decision task on 60 words and 30 nonwords. In a subsequent unexpected memory test condition, during which brain imaging was carried out, the 60 words from the study task were presented intermingled with 30 new words. Subjects were asked to classify each of these 90 words as R (the subject consciously recollected having seen the word in the study phase), K (the subject knew that the word had been in the study phase but did not recollect having seen it then) or N (the subject considered that the word was new, i.e., had not occurred in the study phase).

What is important here are the functional neuroimaging results on correct R trials and on correct K trials (i.e., trials where the subject responded either R or K and the word had indeed been presented at study). What different predictions are made by T_a and T_b about the neuroimaging data from these two types of trial? Henson et al. (1999) do not discuss their data in this way, and do not present their work as attempting to distinguish between dual-process and single-process theory. Instead, the conclusions they state in their paper are about the neural substrates of memory processes, not about psychological theories of memory; that is, their conclusions concerned localisation of function.

However, Henson (2005) does propose to use the data of Henson et al. (1999) to draw conclusions about psychological theories of memory. He argues that T_b (single-process theory) predicts that patterns of brain activity seen on correct K trials will be identical to patterns seen on correct R trials whereas T_a (dual-process theory) predicts that patterns of brain activity seen on correct K trials will be different from patterns seen on correct R trials. Why? Because according to T_a a cognitive process occurs on R trials which does not occur on K trials, namely, recollection, so that there will be brain activity in the brain region responsible for recollection on R trials but not on K trials; hence the two types of trial will produce nonidentical patterns of brain activity. In contrast, T_b asserts that cognition is identical on K and R trials, so patterns of brain activity must also be identical.

But according to Henson et al. (2000) T_b does *not* assert this identity of cognition on the two trial types. According to T_b, on R trials the memory trace is strong, and on K trials it is weak. Henson et al. (2000) propose that when the memory trace is present but weak, subjects, uncertain whether they should respond "Old " or "New", engage in a process of *retrieval monitoring* which is not engaged when the memory trace is strong and the subjects therefore sure that they should respond "Old". Thus according to the one-process theory T_b, retrieval monitoring is a cognitive process (and hence a brain process) which occurs on K trials but not on R trials. Hence T_b, just like T_a, predicts that neuroimaging experiments will show different patterns of brain activation on R trials compared to K trials. It follows that the neuroimaging work of Henson et al. (1999) does not offer a clear example

of a neuroimaging study which could successfully distinguish between two different psychological theories (since the two theories make the same prediction), and indeed Henson et al (2000) acknowledge just this: "While we cannot fully distinguish an explanation in terms of recollection[2] (Yonelinas et al., 1996) from one in terms of high memory strengths or familiarity levels[3] (Donaldson, 1996)…" (p. 918).

Example 2 – Unexpected Memory Testing

In the *subsequent-memory* paradigm, subjects perform some simple task on a series of stimuli and subsequently are presented with a surprise memory test. Henson (2005) described two types of theory that have been proposed to explain results obtained with this unexpected memory test paradigm.

T_a "structural theories" (Cohen and Squire, 1980; Schacter and Tulving, 1994): such theories assert that there exists a cognitive subsystem subserving episodic memory.

T_b "proceduralist" theories (Kolers and Roediger, 1984; Morris et al., 1977): according to this approach, memory as assessed by the unexpected memory test paradigm is better viewed as a byproduct of the processes performed when a stimulus is initially encountered in this paradigm, rather than as reflecting the operation of a distinct memory subsystem (the episodic memory system).

An important difference between these two types of theories is whether successful remembering always involves the occurrence of a specific psychological process (supported by a specialized memory system) when the subsequently-remembered item is initially presented in the study task (this is what is proposed by T_a), or whether successful remembering can be associated with different processes at the time of study if the study task is different in different study conditions (this is what is proposed by T_b).

If subjects' brains are imaged when performing the study task, then each subject's study-phase images can sorted into those images for stimuli that were subsequently remembered and those images for stimuli that were not subsequently remembered. If in addition more than one type of study task is used, one can investigate whether there is any brain region or region in which activation is consistently predictive of subsequent successful recall independently of what the study task was. According to T_a this will be observed. According to T_b it will not.

Otten et al. (2002) carried out such an experiment. Their two study tasks were (a) animate/inanimate classification with printed words or (b) deciding whether the first and last letters in the printed word were in alphabetical order or not.

In both study tasks, activity in left hippocampus correlated with subsequent memory, supporting the structural theory T_a.

However, a subsequent experiment by Otten and Rugg (2001) found clear evidence for the procedural theory, i.e., T_b. This study compared a semantic study task with a phonological study task. Activation of a region within anterior medial prefrontal cortex was associated with better subsequent memory under the semantic task, while activation of regions within left intraparietal and superior occipital cortices were associated with better subsequent memory under the phonological task.

Hence this subsequent-memory research does not provide an example of neuroimaging work successfully distinguishing between two different psychological theories, since the two competing theories receive equal support (or equal lack of support) from the relevant neuroimaging studies.

Example 3 – Tests of Inattentional Blindness *versus* Inattentional Amnesia (Rees et al., 1999)

In a study by Mack and Rock (1998), subjects were asked on each trial to inspect a cross (presented for 200 msec and then backward-masked) and to judge whether its vertical line was longer or shorter than its horizontal line. After three or four such trials, a critical trial was presented: on this critical trial, a small square was presented simultaneously with the cross and within one of the quadrants of the cross. Subjects were asked immediately after this trial whether they had noticed anything different about the trial; 25% of subjects said they had not. Mack and Rock (1998) called this phenomenon "inattentional blindness", proposing that "*there is no perception without attention*… It is essential to bear in mind at the outset that the term *perception* here refers to explicit conscious awareness and is to be distinguished from what is referred to as subliminal, unconscious, or implicit *perception,* that is, perception without awareness. Thus the hypothesis that we believe the evidence presented in this book supports is that there is no conscious perception without attention" (p. 14).

Wolfe (1999) offered a different interpretation of this phenomenon: "Unattended visual stimuli may be seen, but will be instantly forgotten; hence, *inattentional amnesia*" (p. 75). The idea here is that there *is* conscious perception without attention; but there is no *memory* without attention.

In relation to these two theories of the Mack-Rock effect, Rees et al. (1999) commented that "people often are unable to report the content of

[2]That is, T_a.
[3]That is, T_b.

ignored information, but it is unknown whether this reflects a complete failure to perceive it (inattentional blindness) or merely that it is rapidly forgotten (inattentional amnesia)" (p. 2504). Rees et al. (1999) report a brain-imaging study intended to distinguish between two theories about the fate of unattended stimuli. These were:

1) T_a "inattentional blindness": unattended visual stimuli are not perceived.

2) T_b "inattentional amnesia": unattended visual stimuli are perceived, but are then forgotten before they can be reported.

Each visual display used in the experiment by Rees et al. (1999) consisted of a line drawing of an object, with a printed word or random consonant string superimposed on it. The displays were brief (250 msec) and subjects observed a rapid sequence of such displays in epochs lasting 36.9 seconds, during which brain imaging was carried out. The subjects' task was to detect display repetitions, either of pictures (in this condition the letter strings were irrelevant) or of letter strings (in this condition the pictures were irrelevant).

This study relied on previous work on brain imaging of reading in which brain images generated as readers saw random consonant strings were subtracted from brain images generated as readers saw words. This subtraction reveals specific brain regions that are more activated by words than by random letter strings.

In the Rees et al. (1999) study, when letter-string repetition detection was the required task, brain images from 36.9-second epochs when every letter string was a random consonant string were subtracted from images from epochs where 60% of the letter strings were concrete nouns, so as to identify regions of the brain which were more activated by concrete nouns than by random consonant strings. Four such regions emerged: left inferior frontal (BA 44), left posterior temporal (BA 37), left posterior parietal (BA 7/40) and right posterior parietal (BA 7).

In contrast, when picture repetition detection was the required task, no brain regions showed greater activation by words than by letter strings. Rees et al. (1999) argued from this that "the data suggest that word processing is not merely modulated but is abolished when attention is fully withdrawn" (p. 2506) and so offered their data as evidence for T_a and against T_b : "these results demonstrate true inattentional blindness" (p. 2504).

I have three points to make here.

1. *The Picture-Word Interference effect*. When a subject's task is to name pictures, and a task-irrelevant letter string is presented superimposed on each picture, picture naming latency is slower when the superimposed letter string is a word semantically related to the picture than when it is a pronounceable nonword (Rosinski et al., 1975). This and numerous other influences on picture naming exerted by irrelevant letters strings superimposed on the to-be-named pictures have been extensively documented over the past thirty years. If subjects are capable of fully withdrawing attention from printed words so that the words are not processed at all, why don't they do so in the picture-word interference paradigm, since to do so would improve their performance? Rees et al. (1999) do not mention the picture-word interference literature and so do not address this seemingly rather important point.

They do however suggest "that unattended words might be processed to a greater extent under conditions that impose a lower load than the present demanding picture task" (Rees et al., 1999, p. 2506). They might therefore want to suggest that picture naming with superimposed letter strings is a less demanding task than is their picture repetition detection task and to claim that this is why superimposed words get some processing in the picture-naming task but none in the picture repetition detection task. But this would clearly be a circular argument, since the only evidence they could offer that their repetition detection task is the more demanding task is their view that superimposed irrelevant words are not processed when this task is being performed.

Furthermore, even if the circularity of this argument re task demands were overlooked, the argument misses the point concerning why, if subjects can fully withdraw attention from words, they do not do so even when it would be to their advantage (as it would be in the picture-word interference paradigm).

2. *The so-called* "*Visual Word Form Area*". Because Rees et al. (1999) wished to show that the process of visual word recognition is not automatic, but on the contrary can be voluntarily abolished, they needed to show that a region of the brain specific to visual word recognition is inactive in response to the presentation of printed words when the task subjects are performing is picture repetition detection. But they did not show this, because they could not.

The reason that they could not is that no regions of the brain specific to visual word recognition have been found; and that includes what has actually been named the Visual Word Form Area. "Cohen et al. proposed that the Visual Word Form Area (VWFA) in the left mid fusiform gyrus, contrary to its name, is limited to the extraction of an abstract letter string and not involved in proper word recognition" (Kronbichler et al., 2004, p. 946). Regions of the brain that are more activated by printed words than by random letter strings are not more activated by printed words than by printed orthographically legal nonwords (see e.g., Dehaene et al., 2002), so these regions are not selective for words, but at most selective for orthographically legal letter strings, regardless of whether these are words or not.

In order to show that a particular brain region is

selectively sensitive to printed words, it would be necessary to show both:

(a) that this region is more activated by familiar letter strings (i.e., words) than by equally orthographically legal but unfamiliar letter strings (i.e., nonwords);

(b) that this region is more activated by familiar letter strings (i.e., words) than by other equally familiar visual stimuli that are not composed of letters (e.g., familiar faces or objects).

No such brain region or regions have been found. That is very surprising, because it is a trivially easy task for subjects to classify visual stimuli into real words, orthographically legal nonwords, faces and objects. Given this, why has cognitive neuroimaging been unable to detect different brain states associated with these very distinct classes of visual stimuli?

Rees et al. (1999) did acknowledge that "because words differ from consonant strings in several respects (for example, legal orthography and phonology, lexical status, semantics), the activations may involve all the corresponding word-related processes" (p. 2504) but they did not appreciate that this acknowledgement vitiates their conclusions. They wanted to claim that their results showed that subjects "were blind to those properties that distinguish words from random strings of consonants" (p. 2506) but the most they could claim is that their subjects were blind to *one* of those properties (whichever property it is that causes the differential response to words *vs.* random consonants). For example, if subjects are capable of disabling the operation of a brain region that responds more when printed stimuli are orthographically legal than when they are not, that says nothing at all about whether reading is automatic. Such objections could have been avoided had Rees et al. (1999) used orthographically legal nonwords as controls. But then they would not have been able to find brain regions that respond more to words than to nonwords in the letter string repetition detection condition.

3. *What are subjects blind to in inattentional blindness?* Let us suppose that a brain region that has the specific job of visual word recognition had been identified, and that it had then been shown that processing is abolished in that region when attention is not being allocated to visually presented words. That still would not provide an explanation of inattentional blindness, because inattentional blindness is not a matter of visual stimuli *not being processed*, but a matter of such stimuli *not coming to consciousness*, as the quotation from Mack and Rock (1998) above showed.

The regions of the brain that are differentially sensitive to printed words *versus* nonwords are nothing to do with bringing stimuli to consciousness, because they are activated even by stimuli of which the subject does not become conscious (Dehaene et al., 2001). If the subjects studied by Rees et al. (1999) had been inattentionally blind to letter strings in the picture monitoring condition, then, just as Mack and Rock's (1998) subjects were unaware of the presence of the square accompanying the cross on the critical trial, Rees et al.'s (1999) subjects would have been unaware that letter strings had been present in the picture monitoring condition. But Rees et al's (1999) subjects *were* aware of this: "The phenomenal experience when performing the task is indeed of knowing that both a (red) picture and a (green) letter string are present concurrently in the two rapid streams" (p. 2506). Thus these imaging results are not relevant to the psychological phenomenon of inattentional blindness as it was studied by Mack and Rock (1998), by Wolfe (1999) and by others, because the subjects of Rees and colleagues were not inattentionally blind to any visual stimuli.

Thus it is very difficult to accept the conclusion by Rees et al. (1999) that when attention is not paid to a visually presented word cognitive processing of that word is abolished, given the evidence from picture-word interference experiment; but even if this conclusion were correct it would not justify the claim that the inattentional blindness account of the Mack-Rock phenomenon is to be preferred over the inattentional amnesia account, because subjects in the imaging study were not inattentionally blind to the irrelevant words; they were aware that these words were present

Finally, it is perhaps worth noting that the view that reading is automatic, the view which Rees et al. (1999) consider that their neuroimaging data refuted, is currently still widely advocated within the cognitive neuroimaging community: "reading is an obligatory response to visual words" (Kronbichler et al., 2004, p. 946); "computations in the VWFA are largely automatic" (McCandliss et al., 2003, p. 296).

Example 4 – Dependent *versus* Independent Routes for Processing Facial Identity and Facial Expression (Winston et al., 2004)

The aim of this study was not to distinguish between competing psychological theories. On the contrary, it assumed the correctness of a particular psychological theory, the Bruce and Young (1986) theory of face perception. In this theory there are separate modules for representing the identity of a known face and for processing changeable aspects of faces such as gaze direction and facial expression. The hypothesis tested in this imaging study was that the first of these modules is located in fusiform cortex and that second module is located in superior temporal sulcus. So this was a

neuroanatomical localisation study (rather than a study evaluating competing psychological theories) and hence outside the scope of this paper (and, for the same reason, also outside the scope of the paper by Henson, 2005).

Example 5 – Verbal *versus* Visuospatial Slave Systems in Working Memory (Smith and Jonides, 1997)

Exactly the same point can be made re this paper as made above re the Winston et al. (2004) paper. Smith and Jonides (1997) did not apply neuroimaging data to the task of distinguishing between competing psychological theories: on the contrary, they assumed a particular psychological theory (the theory of working memory proposed by Baddeley, 1986, 1992) and carried out neuroimaging experiments which sought to discover the neuroanatomical loci of particular nodules posited by this theory. The object of this work was not to support or challenge any particular psychological theory and the conclusions drawn from the data almost entirely concern localisation of function.

Smith and Jonides (1997) do assert "Our results provide support for certain cognitive models of working memory (e.g., Baddeley, 1992)" but their assertion is not justified, for two reasons.

Firstly, alternative models are not considered; the results might have been consistent with alternatives to the Baddeley (1992) model, in which case they could not have been offered as support for the Baddeley model.

Secondly, what pattern of results could Smith and Jonides (1997) have obtained which would have *challenged* the Baddeley model? A complete failure to localise in the brain any of the modules of the model would not have challenged the Baddeley model, since that model does not claim that its modules are localised in discrete brain regions. If there is no possible pattern of results that would have challenged the model, then no particular pattern of results can be offered as supporting the model.

Example 6 – "Simulation Theory" *versus* "Theory Theory" of Representing Others' Intentions (Ramnani and Miall, 2004)

It is generally agreed that the way I predict what another person is going to do is by representing that person's current mental states in my own mind. Given this, how do I use such representations to predict the other person's behaviour?

Two theories have been proposed here.

1) T_a "simulation theory": I simulate in my own mind the other person's mental states; the intentions that come to my own mind as I run this simulation I can take as representing what the other person intends to do, and so as predicting what that other person will do.

2) T_b "theory theory": rather than directly reading off the results of a simulation of the other person's mind that I run on my own mind, I create a theory about the other person's mind, and seek to deduce from that theory what actions would be taken by a person with a mind like that.

Ramnani and Miall (2004) are explicit about the aims of their neuroimaging study: "The problem of understanding others, intentions can be translated into the more tangible problem of predicting others, actions. Identifying the neural mechanisms used to predict the actions of others may… enable us to distinguish between the two theories" (p. 85).

Subjects' brains were scanned as they prepared their own actions or predicted the future actions of other people. When a person is preparing to execute an action, that person's premotor cortex is activated. If the way we predict other people's actions is by simulating them, then our premotor cortex should also be activated by the task of predicting others' actions. So simulation theory predicts that, in the condition in which a subject is predicting how another person will act, there will be activation of that subject's premotor cortex. Theory theory does not predict activation of premotor cortex in this prediction condition. Instead, theory theory predicts activation of the paracingulate cortex and the superior temporal sulcus, because these are "areas most consistently activated when subjects evaluate the intentions of others" (Ramnani and Miall, 2004, p. 85).

The results, however, did not turn out to favour just one of the theories. In the predict-companion condition, motor cortex *was* activated. That is consistent with T_a and inconsistent with T_b. However, the region of premotor cortex which was activated in the predict-companion condition was different from the region activated when subjects are preparing their own actions. And what is more, in this predict-companion condition paracingulate cortex and superior temporal cortex were activated; that is consistent with T_b and inconsistent with T_a. Hence no unequivocal support for either of these psychological theories was obtained. The Abstract concludes: "We provide compelling evidence that areas within the action control system of the human brain are indeed activated when predicting others' actions" (this, as I have noted, is evidence against T_b) "but a different action sub-system is activated when preparing one's own actions" (this, as I have noted, is evidence against T_a). Consequently, the study by Ramnani and Miall (2004) does not provide an example in which cognitive neuroimaging data were successfully used to distinguish between competing psychological theories.

Conclusions

Rather a lot of people believe that you can't learn anything about cognition from studying the brain.

"The essence of these arguments is that distinct neurological structures need not correspond to functional modules – indeed, there might not be any modules. To be able to decide whether there are and whether there is any correspondence, you need to have a complete theory of cognition before you begin interpreting images. Hence imaging can, in principle, add nothing new. There is a level of psychological theorising – the cognitive level – that can only be studied at this level, and information from lower levels will tell us nothing about what happens at the cognitive level. In summary, we need a theory of cognition before we can properly understand what is happening at the lower levels" (Harley, 2004, pp. 10-11).

"The other possible aim of cognitive neuroimaging is to use imaging data for testing or adjudicating between cognitive models. Here the ultra-cognitive-neuropsychological position is a particularly extreme one. The assertion is that this aim is impossible to achieve in principle, because facts about the brain do not constrain the possible natures of mental information-processing systems. No amount of knowledge about the hardware of a computer will tell you anything serious about the nature of the software that the computer runs. In the same way, no facts about the activity of the brain could be used to confirm or refute some information-processing model of cognition" (Coltheart, 2004, p. 22).

"Functional neuroimages do not constrain cognitive models of language processing" (Paap, 1997).

"To mix hardware and program descriptions ('that transistor has a missing right parenthesis') is to make a category mistake. The conceptual distance between symbolic rules and neurons is so great that it is difficult to propose how knowledge about one might contribute to knowledge of the other" (Colby, 1978).

"… The question of how the elements of the model are implemented in the brain . . the question of what functions are performed in particular anatomical locations. I believe these are important questions, but the answers do not affect the form or the nature of the psychological model" (Morton, 1984, pp. 40-41).

"Neuroimages do not isolate modular components that correspond to cognitive modules" (van Orden and Paap, 1997, p. 93).

"Even if we could associate precisely defined cognitive functions in particular areas of the brain (and this seems highly unlikely), it would tell us very little if anything about how the brain computes, represents, encodes, or instantiates psychological processes" (Uttal, 2001, p. 217).

"I once gave a (perfectly awful) cognitive science lecture at a major centre for brain imaging research. The main project there, as best I could tell, was to provide subjects with some or other experimental tasks to do and take pictures of their brains while they did them. The lecture was followed by the usual mildly boozy dinner, over which professional inhibitions relaxed a bit. I kept asking, as politely as I could manage, how the neuroscientists decided which experimental tasks it would be interesting to make brain maps for. I kept getting the impression that they didn't much care. Their idea was apparently that experimental data are, *ipso facto*, a good thing; and that experimental data about when and where the brain lights up are, *ipso facto*, a better thing than most. I guess I must have been unsubtle in pressing my question because, at a pause in the conversation, one of my hosts rounded on me. 'You think we're wasting our time, don't you?' he asked. I admit I didn't know quite what to say. I've been wondering about it ever since" (Fodor, 1999).

How might these people be shown the error of their ways? All that is needed to do this is to provide them with actual examples where neuroimaging data have successfully been used to distinguish between competing psychological theories. They all claim that this cannot happen. Has it ever happened?

Acknowledgements. I thank Rik Henson for much stimulating discussion.

References

Baddeley AD. *Working Memory.* Oxford: Clarendon Press, 1986.

Baddeley AD. Is working memory working? The Fifteenth Bartlett Lecture. *Quarterly Journal of Experimental Psychology, 44:* 1-31, 1992.

Blazely A, Coltheart M and Casey B. Semantic dementia with and without surface dyslexia. *Cognitive Neuropsychology, 22:* 695-717, 2005.

Bruce V and Young A. Understanding face recognition. *British Journal of Psychology, 77:* 305-327, 1986.

Cipolotti L and Warrington EK. Semantic memory and reading abilities: A case report. *Journal of the International Neuropsychological Society, 1:* 104-110, 1995.

Cohen NJ and Squire LR. Preserved learning and retention of pattern analysing skills in amnesia. *Science, 210:* 207-209,1980.

Colby KM. Mind models: An overview of current work. *Mathematical Biosciences, 39:* 159-195, 1978.

Coltheart M. Lexical access in simple reading tasks. In Underwood G (Ed), *Strategies of Information Processing.* London: Academic Press, 1978.

Coltheart M. Brain imaging, connectionism and cognitive neuropsychology. *Cognitive Neuropsychology, 2:* 21-25, 2004.

Coltheart M, Curtis B, Atkins P and Haller M. Models of reading aloud: Dual-route and parallel-distributed-processing approaches. *Psychological Review, 100:* 589-608. 1993.

Coltheart M, Rastle K, Perry C, Langdon R and Ziegler JC. DRC: A dual route cascaded model of visual word recognition and reading aloud. *Psychological Review, 108:* 204-256, 2001

Dehaene S, Le Clec HG, Poline JB, Le Bihan D and Cohen L. The visual word form area: A prelexical representation of visual words in the fusiform gyrus. *NeuroReport, 13:* 321-325, 2002

Dehaene S, Naccache L, Cohen L, Lebihan D, Mangin JF, Poline J-B and Rivière D. Cerebral mechanisms of word masking and unconscious repetition priming. *Nature Neuroscience, 4:* 752-758, 2001

DONALDSON W. The role of decision processes in remembering and knowing. *Memory and Cognition*, *24:* 523-533, 1966.
FODOR J. Let your brain alone. *London Review of Books, 21:* 1999.
GERHAND S. Routes to reading: A report of a non-semantic reader with equivalent performance on regular and exception words. *Neuropsychologia, 39:* 1473-1484, 2001.
GOODALL WC and PHILLIPS WA. Three routes to reading aloud: Evidence from a case of acquired dyslexia. *Cognitive Neuropsychology, 12:* 113-147, 1995
HARLEY TA. Does cognitive neuropsychology have a future? *Cognitive Neuropsychology, 21:* 2-16, 2004
HARM MW and SEIDENBERG MS. Phonology, reading acquisition and dyslexia: Insights from connectionist models. *Psychological Review, 106:* 491-528, 1999
HARM MW and SEIDENBERG MS. Computing the meaning of words in reading: Cooperative division of labor between visual and phonological processes. *Psychological Review, 111:* 662-672, 2004.
HEATHCOTE A. Item recognition memory and the receiver operating characteristic. *Journal of Experimental Psychology: Learning, Memory and Cognition, 29:* 1210-1230, 2003.
HENSON R. What can functional neuroimaging tell the experimental psychologist? *Quarterly Journal of Experimental Psychology, 58A:* 193-233, 2005.
HENSON RN, RUGG MD, SHALLICE T, JOSEPHS O and DOLAN RJ. Recollection and familiarity in recognition memory: An event-related fMRI study. *Journal of Cognitive Neuroscience, 19:* 3962-3972, 1999.
HENSON RN, RUGG MD, SHALLICE T, JOSEPHS O and DOLAN RJ. Confidence in recognition memory for words: Dissociating right prefrontal cortex in episodic retrieval. *Journal of Cognitive Neuroscience, 12:* 913-923, 2000.
KOLERS PA and ROEDIGER HL. Procedures of mind. *Journal of Verbal Learning and Verbal Behaviour, 23:* 425-449, 1984.
KRONBICHLER M, HUTZLER F, WIMMER H, MAIR A, STAFFEN W and LADURNER G. The visual word form area and the frequency with which words are encountered: Evidence from a parametric fMRI study. *NeuroImage, 21:* 946-953, 2004.
LAMBON-RALPH MA, ELLIS A and FRANKLIN S. Semantic loss without surface dyslexia. *Neurocase, 1:* 363-369, 1995.
LYTTON WW and BRUST JC. Direct dyslexia: Preserved oral reading of real words in Wernicke's aphasia. *Brain, 112:* 583-594, 1989.
MACK A and ROCK I. *Inattentional Blindness.* Cambridge, MA: MIT Press, 1998.
MCCANDLISS BD, COHEN L and DEHAENE S. The visual word form area: Expertise for reading in the fusiform gyrus. *Trends in Cognitive Sciences, 7:* 293-299, 2003.
MORRIS CD, BRANSFORD JD and FRANKS JJ. Levels of processing versus transfer appropriate processing. *Journal of Verbal Learning and Verbal Behavior, 16:* 519-533, 1977.
MORTON J. Brain-based and non-brain-based models of language. In Caplan D (Ed), *Biological Perspectives on Language.* Cambridge, MA: MIT Press, 1984.
OTTEN LJ, HENSON R and RUGG MD. State-related and item-related neural correlates of successful memory encoding. *Nature Neuroscience, 5:* 1339-1344, 2002.
OTTEN LJ and RUGG MD. Task-dependency of the neural correlates of episodic encoding as measured by fMRI. *Cerebral Cortex, 11:* 1150-1160, 2001.
PAAP KR. Functional neuroimages do not constrain cognitive models of language processing. Paper presented at the 22nd Annual Interdisciplinary Conference, Jackson Hole, Wyoming, Feb 4 1997.
PATTERSON K and SHEWELL C. Speak and spell: Dissociations and word-class effects. In Coltheart M, Sartori G and Job R (Eds), *The Cognitive Neuropsychology of Language*. Hove: Lawrence Erlbaum Associates, 1987.
PLAUT DC, MCCLELLAND JL, SEIDENBERG MS and PATTERSON KE. Understanding normal and impaired word reading: Computational principles in quasi-regular domains. *Psychological Review, 103:* 56-115, 1996.
POEPPEL D. A critical review of PET studies of phonological processing. *Brain and Language, 55:* 317-351, 1996.
RAMNANI N and MIALL RC. A circuit in the human brain for predicting the actions of others. *Nature Neuroscience, 7:* 85-90, 2004.
RASTLE K and COLTHEART M. Is there serial processing in the reading system, and are there local representations? In Andrews S (Ed), *All about Words: Current Issues in Lexical Processing*. Hove: Psychology Press, in press.
REES G, RUSSELL C, FRITH C and DRIVER J. Inattentional blindness versus inattentional amnesia for fixated but ignored words. *Science, 286:* 2504-2506, 1999.
ROGERS TT, LAMBON RALPH MA, HODGES JR and PATTERSON K. Natural selection: The impact of semantic impairment on lexical and object decision. *Cognitive Neuropsychology, 11:* 331-352, 2004.
ROSINSKI RR, GOLINKOFF RM and KUKISH KS. Automatic semantic processing in a picture-word interference task. *Child Development, 46:* 247-253, 1975.
SCHACTER DL and TULVING E. *Memory Systems.* Cambridge, MA: MIT Press, 1994.
SEIDENBERG MS and MCCLELLAND JL. A distributed, developmental model of word recognition and naming. *Psychological Review, 96:* 523-568, 1989.
SMITH EE and JONIDES J. Working memory: A view from neuroimaging. *Cognitive Psychology, 33:* 5-42, 1997
UTTAL WR. *The New Phrenology: The Limits of Localizing Cognitive Processes.* Cambridge, MA: MIT Press, 2001.
VAN ORDEN GC and PAAP KR. Functional neuroimages fail to discover pieces of mind in parts of the brain. *Philosophy of Science Proceedings, 64:* S85-S94, 1997.
WINSTON JS, HENSON RNA, FINE-GOULDEN MR and DOLAN RJ. FMRI-adaptation reveals dissociable neural representations of identity and expression in face perception. *Journal of Neurophysiology, 92:* 1830-1839, 2004.
WOLFE J. Inattentional amnesia. In Coltheart V (Ed), *Fleeting Memories.* Cambridge, MA: MIT Press, 1999.
YONELINAS AP. The nature of recollection and familiarity: A review of 30 years of research. *Journal of Memory and Language, 46:* 441-517, 2002.

Max Coltheart, Macquarie Centre for Cognitive Science, Macquarie University, Sydney, NSW, Australia 2109. e-mail: max@maccs.mq.edu.au