

## PERHAPS FUNCTIONAL NEUROIMAGING HAS NOT TOLD US ANYTHING ABOUT THE MIND (SO FAR)

Max Coltheart

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

The issue with which this Forum is concerned has been very nicely illustrated in a recent article by Chomsky (2000). He is discussing the computer model of the mind, using Block (1990) as a standard exposition of this position, and he has this to say (the internal quotes are all from Block's chapter):

"According to the computer model of the mind cognitive science 'aims for a level of description of the mind that abstracts away from the biological realizations of cognitive structures'. It does so in principle, not because of a lack of understanding we hope will be temporary, or to solve some problem for which implementation is irrelevant, or in order to explore the consequences of certain assumptions. Rather, for cognitive science 'it does not matter' whether one chooses an implementation 'in grey matter... switches or cats and mice'. If we can construct automata in 'our *computational image*' performing as we do by some criterion, then 'we will naturally feel that the most compelling theory of the mind is one that is general enough to apply to both them and us'. For the computer model of the mind [it] follows that nothing discovered about the brain will matter for the cognitive sciences. For example, if it is some day discovered that one interpretation of the recursive procedure can be implemented at the cellular level, and another cannot, the result will be irrelevant to the study of the human language. That does not seem to me to be a wise course" (Chomsky, 2000, pp. 21-22).

Here Chomsky (2000), having described the computer model of the mind well, misunderstands its position re the relevance of facts about the brain. That approach to understanding cognition does *not* assert that if some fact about the brain were discovered which showed that one type of cognitive operation could be implemented in the brain and another could not, that would be irrelevant to the study of human cognition. What that approach asserts is quite different, namely, that there are no such facts to discover "... 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).

Suppose we have two theories  $T_a$  and  $T_b$  about how recursion in language is achieved by human speaker/hearers. Suppose we could indeed show that one could be implemented in the human brain whilst the other could not. Should that lead cognitive scientists to adopt  $T_0$  and reject  $T_1$ ? Surely all would answer "yes"? The issue is not what ought to occur if this happened; it is whether it ever could happen. Block (1990), Morton (1984) and others suggest that it could not, and *that* is why "for the computer model of the mind [it] follows that nothing discovered about the brain will matter for the cognitive sciences". So the question is: could there be facts about the brain which are relevant to evaluating psychological theories (such as theories about how recursion in language processing is dealt with)? Possible answers to this question are:

a) Yes, and here are some of them (this appears to be the view of some contributors to this Forum);

b) Not right now, but one day there could be (this appears to be Chomsky's view in the case of theories about recursion);

c) No (this appears to be the view of Page, 2006, this issue, and of various authors quoted at the end of the lead article in this Forum).

This Forum was of course restricted to considering only one source of data about the brain, namely, functional imaging; and it was also deliberately restricted to what has happened so far in the existing literature. What might be possible in principle in the future was beyond the scope of the Forum (but not beyond the scope of the article by Page, 2006, this issue).

On the whole, contributors to the Forum believed that they could offer examples of type (a); but I have not been persuaded by these, and so must explain my reasons for not having been persuaded.

Let's first consider Umiltà (2006, this issue), whose domain of cognition was visual attention. He accepted the task of seeking to show that brain imaging data have allowed one to distinguish between two competing psychological theories of visual attention. It is important to scrutinise his

article to find out exactly how these two theories are characterized. I think this is a fair summary:

$T_a$ : endogenous and exogenous attention are governed by a single cognitive system

$T_b$ : endogenous and exogenous attention are governed by separate cognitive systems

He argues that imaging has revealed that endogenous attention activates a dorsal parietofrontal network whereas exogenous attention activates a ventral parietofrontal network, and takes this as evidence that supports  $T_b$  and not  $T_a$ . However, he says nothing about the nature of these two theories at the psychological or cognitive level: nothing for example about the different functional architectures that the two theories propose<sup>1</sup>; indeed, as he says, “It is not always easy to determine if a theory is psychological or not” (p. 400). I think one can show that the two theories he considers are not psychological because nothing in his paper would be changed if he had stated the two theories thus:

$T_a$ : endogenous and exogenous attention are governed by a single brain system

$T_b$ : endogenous and exogenous attention are governed by separate brain systems

These are theories about the brain, not about the cognitive level.

What’s more, Umiltà (2006, this issue) himself acknowledges: “The index of true exogenous spatial orienting would be the presence of inhibition. Inhibition, however, cannot yet be measured in functional neuroimaging studies” (p. 400); and his paper concludes with additional doubts about what functional neuroimaging has really told us about psychological theories of spatial attention.

Jonides et al. (2006, this issue) are much more sanguine about the contributions of functional neuroimaging to cognitive psychology – “so many examples, so little space”, they say in their title. They offer the following competing theories of working memory:

1)  $T_a$ : working memory for verbal information is mediated by a different cognitive system than working memory for spatial information (the Baddeley-Hitch model)

2)  $T_b$ : there is a single cognitive system of storage and rehearsal processes that works on both verbal and spatial information

Now, this raises a point which I ought to have made in the lead article but omitted to. Suppose I claim that there cannot be any facts about the brain that could conflict with any cognitive theory expressed at the psychological level. You could reply: someone proposing a model of memory in which some memory store was postulated to have infinite capacity would surely be refuted by one particular fact about the brain, namely, that it turns

out to be finite in size. Of course that refutes my general claim; but not in any interesting way, because the psychological theory involved here is not one that has ever been seriously entertained. So the claim needs to be reformulated to exclude such degenerate cases: it could go something like “There cannot be any facts about the brain that could conflict with any otherwise plausible cognitive theory expressed at the psychological level”. What’s the criterion for “otherwise plausible”? A reasonable criterion is that there be psychologists who, at the time of collection of the imaging data, were supporters of the theory in question. Now we can return to the example proposed by Jonides et al. (2006, this issue). When they did their functional neuroimaging study in the mid 1990s,  $T_a$  had many supporters – it is, after all part of the Baddeley-Hitch model of working memory. But at that time who was rejecting  $T_a$  and arguing instead that  $T_b$  was correct? No one, as far as I know. Thus I can’t agree that this imaging study can really be seen as an attempt to distinguish between two psychological theories.

But suppose that there really had been a viable  $T_b$ . Could the results have then be taken as supporting  $T_a$ ? Jonides et al. (2006, this issue) ignored the point I made in the lead article here, a point expanded upon in the following article by Page (2006, this issue). The finding that different parts of the brain are associated with visual and verbal working memory is *not* support for  $T_a$  because  $T_a$  does not predict this result. That’s easy to demonstrate. Suppose the imaging study had found that the same brain regions were involved in visual and verbal working memory: would that be inconsistent with  $T_a$ ? No. So all possible results of this study are compatible with  $T_a$  and so no result could have contradicted this theory. I think the same is true for the example offered by Vallar (2006, this issue) concerning rehearsal and speech production: the theory that the process of rehearsal is cognitively independent of the processes of speech production does not predict that different regions of the brain will be activated by these processes. But here I am on uncertain ground because I have not understood what Vallar’s (2006, this issue) models  $T_a$  and  $T_b$  are claiming, nor what the actual imaging results are which he has brought to bear on these two models.

However, there is a pair of theories between which the results of Smith and Jonides (1997) *did* adjudicate, namely

$T_a$ : working memory for verbal information is mediated by a different brain system than working memory for spatial information.

$T_b$ : there is a single brain system of storage and rehearsal processes that works on both verbal and spatial information.

The functional neuroimaging results supported  $T_a$  and conflicted with  $T_b$ . But of course  $T_a$  and  $T_b$  here are not psychological theories: they are

<sup>1</sup>Exactly the same point applies to Umiltà’s (2006, this issue) discussion of visuospatial neglect.

theories about the brain, and so not relevant to this Forum.

Jonides et al. (2006, this issue) provide three other examples, one to do with interference-resolution, one to do with adapted-item and verb-generation tasks, and one to do with visual imagery.

I'll consider first the example involving interference-resolution. The competing theories here are:

1)  $T_a$ : there is only one type of conflict that arises from irrelevant competing information.

2)  $T_b$ : conflict arising from previous information in working memory (proactive interference) is "separable" from conflict that results from previous responses.

Whether these are psychological theories depends completely on what is meant by "separable". It turns out that "separable" means "neurally separable" here, because the conclusion of the relevant neuroimaging study is reported thus: "The dissociation of these two regions suggests that interference-resolution from conflicting working memory representations is mediated by different neural structures than those resulting from response-conflict". The authors seek to infer something psychological from this result: "This supports the idea that interference-resolution is not a unitary construct, as the behavioural data might imply, but rather separate forms of resolution may act upon different forms of conflict". This might allow a reformulation of  $T_b$  that is closer to being psychological, something like:

$T_b$ : the form of resolution used to resolve conflict arising from previous information in working memory (proactive interference) is different from the form of resolution used to resolve conflict that results from previous responses.

But since we are told nothing about what these forms of resolution might be – e.g. about the different functional architectures of the two conflict-resolution systems – there isn't really any psychological theory here.

I don't need to discuss the example using adapted-item and verb-generation tasks, because Jonides et al. (2006, this issue), discussing the neuroimaging results here, observe that "Jonides et al. (2006) discuss various alternative psychological accounts that might be consistent with these facts" (p. 416). So here the example is one where neuroimaging results specifically did *not* distinguish between psychological theories.

Finally, the visual imagery debate, expressed by Jonides et al. (p. 416) as follows:

$T_a$ : mental images have no special status: they are just one example of propositional mental representations similar to ones that underlie natural language.

$T_b$ : mental imagery requires a different class of processes than those used in language, a class that makes heavy use of processes used to perceive in the face of sensory stimulation.

The key issue here is whether the brain mechanisms activated when we are seeing are the brain mechanisms activated during visual imagery. According to Jonides et al. (2006, this issue), in the neuroimaging work of Ganis et al. (2004) on this topic, although there were similarities in brain activation in perception and imagery, "there were some important differences in brain activation between perception and imagery as well, leading to refined hypotheses about what is shared in common and what distinguishes these two cognitive functions". In other words, this work did not favour one of  $T_a$  and  $T_b$ ; instead the conclusion was that we need a new theory,  $T_c$ .

Caplan and Chen's (2006, this issue) contribution to the Forum reports some interesting behavioural and neuroimaging data concerning sentence comprehension, but acknowledges "Coltheart (2006, this issue) has challenged researchers to produce an example of a study in which neurovascular effects decide between two competing models of a cognitive process. This paper falls short of addressing this challenge" (p. 393).

Cappa (2006, this issue) refers to Fodor thus: "Who cares about localization *per se*? As effectively phrased by Fodor: 'If the mind happens in space at all, it happens somewhere north of the neck. What exactly turns on knowing how far north?' Localization *per se* is not interesting, and from this point of view one may agree with Fodor". I think this misunderstands Fodor. He was not claiming localization was *per se* uninteresting; instead, he was raising the issue of what questions of interest to cognitive science (questions about theories expressed at the cognitive level, that is) could be answered by facts about localization (I take it that the answer he had in mind was: none). Thus Fodor was expressing just the view expressed by Block (1990) as quoted by Chomsky (2000) at the beginning of this article. And what is more, despite Cappa's (2006, this issue) negative remarks about localization, all of the examples he gave of neuroimaging research concerned the localization of cognitive functions – none were offered as examples in which neuroimaging data have been or might be used to distinguish between psychological theories. The contribution by Schutter et al. (2006, this issue) also did not offer any such examples; indeed, these authors seem to accept the view that there have been no such examples (so far).

Seron and Fias (2006, this issue), it seems to me, came close to providing the kind of example that I was seeking. They offer not two but three competing theories expressed at the psychological level. These theories are alternative psychological theories of how number transcoding tasks such as reading aloud Arabic numerals are performed. The theories are:

1)  $T_a$ : such transcodings always require passing through a semantic level.

2)  $T_b$ : such transcodings bypass the semantic level so make no use of it.

3)  $T_c$ : such transcodings can make use of both a semantic route and a nonsemantic route (with various factors biasing the use of one route or other).

Amongst the virtues of this proposal is that all three of these theories have current proponents.

The next step is to nominate the intraparietal sulcus (IPS) as a region of the brain that is activated when semantic tasks involving numbers are being performed. Let's accept this nomination, and measure IPS activation when people are performing number transcoding tasks. The predictions seem clear:

$T_a$ : IPS will always be activated when such tasks are being performed.

$T_b$ : IPS will never be activated when such tasks are being performed.

$T_c$ : IPS will sometimes be activated when such tasks are being performed and sometimes not, as a function of the various factors biasing the use of one route or other.

But, as Seron and Fias (2006, this issue) then go on to show, "the story turns out to be more difficult and the Coltheart game not so easy" (p. 407). They point out two crucial obstacles to this way of seeking to adjudicate between the three competing psychological theories.

The first is to do with automaticity. It is a commonplace fact of cognitive psychology that elements of any cognitive system that has been acquired to a high degree of skill will be accessed even when they are not required for performance of a task required of a subject. The Stroop effect shows that words access their lexical representations even when the task (colour-naming of letter strings) does not require this, and indeed even when such access harms performance of the task. Similarly the IPS is activated even in tasks that don't require access to number semantics, and even when numbers are presented briefly and masked so that the subject is not aware of them. So it does not follow from  $T_b$  and  $T_c$  that there will ever be occasions when a numeral does not excite the IPS, even if on all ( $T_b$ ) or some ( $T_c$ ) occasions this is not needed for performing the transcoding task. Thus the finding that the IPS is always activated when subjects are performing a transcoding task is compatible with all three theories, and so cannot be used to distinguish between them. This issue about automaticity is frequently overlooked in the design of imaging studies.

The second crucial obstacle is that activation of the IPS is not coextensive with accessing number semantics; the IPS is activated by various cognitive activities that have nothing to do with numbers. So if the IPS is activated in one number-processing task and not in another, this difference can't be taken as evidence that number semantics is activated in one task and not the other; the

difference might depend on one of the other cognitive activities with which the IPS is associated.

Henson (2006, this issue) makes several points. Firstly, he suggests (p. 387) "that datasets rarely, if ever, provide unequivocal evidence for a theory". The motivation for this gambit is as follows: if someone claims that there are no neuroimaging data that have provided unequivocal evidence for a psychological theory, one defence of cognitive neuroimaging would be to seek to refute this claim by providing such examples; but an alternative response is Henson's Gambit, which is to agree that cognitive neuroimaging offers no such examples (that is the sacrificial element of the gambit), and then to claim that there are no behavioural data either that have provided unequivocal evidence for a psychological theory (thus seeking to recover the sacrificed piece).

It was in anticipation of this move that I offered in the lead article two examples where I believe it is clear that behavioural data *have* provided unequivocal evidence for a theory. Note that I was not claiming, as Vallar (2006, this issue) seems to think I was, that there can be *experimenta crucium* here: single experiments which provide unequivocal evidence for one psychological theory and against another. I offered instead examples where there was a collection of relevant results all of which favoured one theory and were inconsistent with the competitor. "Unequivocal" here means that all the relevant results speak with a single voice i.e. all favour the same theory.

I believe that such examples are easy to find in cognitive psychology and that Henson's criticisms of the two I offered do not stand up to scrutiny. With regard to the example of reading in semantic dementia, Henson misdescribes the prediction from  $T_a$ , which is that *every* patient with semantic dementia will be surface dyslexic; several studies have reported patients with semantic dementia who were not surface dyslexic, so this prediction fails. The fact that only a small minority of patients with semantic dementia are not surface dyslexic is irrelevant:  $T_a$  says that there will be *no* such patients. With regard to the example involving serial processing in the reading system, Rastle and Coltheart (in press) have described the eight different datasets that bear on whether there is serial processing in the reading system, and the verdict here is unequivocal: all eight provide evidence of serial processing, so here there is unequivocal support for a serial processing theory and unequivocal evidence against a purely parallel processing account of reading. So I pursue the chess analogy by offering the view that the present status of Henson's Gambit is that it should be considered as having been refuted.

Henson had space for reconsidering, in response to my comments on Henson (2005), only one of his original examples of cases where he considered



that cognitive neuroimaging had succeed in distinguishing between psychological theories: the example concerning the Remember/Know distinction. In the original discussions, there were two competing theories  $T_a$  and  $T_b$ , expressed purely in psychological terms, and the issue was whether there existed neuroimaging data that distinguished between these two theories. In his commentary, he introduces a third psychological theory  $T_c$  and observes (p 389) that the situation now is that “the data previously explained by one theory is also explained by a new, alternative theory”; but he also refers to new neuroimaging data which he believe favours  $T_a$  over both  $T_b$  and  $T_c$ . It may be that if this argument is spelled out further – by explaining why  $T_a$  actually predicts these new data and explaining why they are inconsistent with  $T_b$  and  $T_c$  – a clear example of neuroimaging constraining cognitive theory might emerge. But I doubt that it will be possible to spell out just such an argument; and any attempt to do so would have to take into account the points concerning the Remember/Know research made by Page (2006, this issue) in his viewpoint article, and also the extensive new results and new analyses of old results provided by Wixted and Stretch (2004) which these authors offer as (behavioural) evidence against  $T_a$ .

Finally, I consider the instructive example offered by Jack et al. (2006, this issue). This concerns visual search and in particular “inefficient visual search” – the form of visual search in which search RT increases with the size of the set to be searched – and offers two competing theories as to the mechanism by which such search is effected.  $T_a$  claims that inefficient visual search is accomplished by a parallel search of the display;  $T_b$  claims that it is accomplished by a serial search of the display. These are clearly theories at a psychological level; they are clearly competing theories; they are explicit about the psychological mechanisms involved; and both theories have their adherents. So this is a good example; and Jack et al. (2006, this issue) assert that there are neuroimaging data which support  $T_b$  and conflict with  $T_a$ .

Their argument runs as follows. If  $T_b$  is correct, then the visual search task requires subjects to covertly shift visual attention to different regions of the visual display. If  $T_a$  is correct, no such covert shifting of visual attention will occur during visual search. The neuroimaging data of Corbetta et al. (1993) show that a region in posterior parietal cortex, especially right posterior parietal cortex, is responsible for the task of covertly shifting visual attention. So if  $T_b$  is correct, this parietal region will be more active when subjects perform inefficient visual search than when they perform efficient visual search, whereas if  $T_a$  is correct there will be no difference here. Corbetta et al. (1995) reported that activation in this region was greater during inefficient than efficient visual search, and this is the imaging result which Jack

et al. (2006, this issue) take as supporting  $T_b$  and conflicting with  $T_a$ .

This reasoning required that covert shifting of visual attention and activation of this right posterior parietal region be *co-extensive*. Two things must be true: whenever there is covert shifting of visual attention there must always be activation of this brain region, and whenever this brain region is active there must always be covert shifting of visual attention going on. Thus the claim is that the sole function of this brain region is control of covert shifting of visual attention; unless that is so, the reasoning about  $T_a$  and  $T_b$  does not follow.

But subsequent work has shown that covert shifting of visual attention is *not* the sole function of this parietal region, because the region is activated when subjects are searching for conjunction targets (compared to when searching for single-feature targets) even when there is never more than one stimulus present and so covert shifting of visual attention will not be going on (Wojciulik and Kanwisher, 1999). The relevant right parietal region is also activated in an orientation and location detection task that does not require any covert shifting of visual attention (Vandenberghe et al., 1997). Thus it turns out not to be the case that “activity in posterior parietal cortex can be used to index the process of shifts in spatial attention” (Jack et al., 2006, this issue, p. 420); and if this is not the case, then the reasoning upon which Jack et al. (2006, this issue) based their example can’t be sustained. This is exactly the same problem identified by Seron and Fias (2006, this issue) in relation to cognitive neuroimaging of number processing: the reasoning only follows if the key brain region associated with number semantics, the intraparietal sulcus, has only this cognitive function and no other, but that is not so: it is activated by other cognitive tasks that have nothing to do with number semantics.

#### ENVOI

A fact sheet currently available from the National Institute of Mental Health (NIMH) website says:

“Neuroimaging in cognitive research. Neuroimaging techniques that allow scientists to peer inside the living, functioning human brain are immensely powerful tools for cognitive research”.

Not all agree with this: the neurosurgeon Henry Perowne (protagonist of Ian McEwan’s novel *Saturday*) laments on p. 243 of the book:

“If only [it were so] that penetrating the skull brings into view not the brain but the mind”.

#### REFERENCES

- BLOCK N. The computer model of the mind. In Osherson D and Smith E (Eds), *Thinking: An Invitation to Cognitive Science* (vol. 3). Cambridge, MA: MIT Press, 1990.

- CAPLAN D. Using fMRI to discover cognitive operations. *Cortex*, 42: 393-395, 2006.
- CAPPA SF. Brain imaging: Useful, helpful, beneficial? *Cortex*, 42: 396-398, 2006.
- CHOMSKY N. Linguistics and brain science. In Marantz A, Miyashita Y and O'Neill W (Eds), *Image, Language, Brain*. Cambridge: MIT Press, 2000.
- CORBETTA M, MIEZIN FM, SHULMAN GL and PETERSON SE. A PET study of visuospatial attention. *Journal of Neuroscience*, 13: 1202-1226, 1993.
- CORBETTA M, SHULMAN GL, MIEZIN FM and PETERSON SE. Superior parietal cortex activation during spatial attention shifts and visual feature conjunction. *Science*, 270: 802-805, 1995.
- GANIS G, THOMPSON WL and KOSSLYN SM. Brain areas underlying visual mental imagery and visual perception: An fMRI study. *Cognitive Brain Research*, 20: 226-241, 2004.
- HENSON R. What can functional neuroimaging tell the experimental psychologist? *Quarterly Journal of Experimental Psychology*, 58A: 193-233, 2005.
- HENSON R. What has (neuro)psychology told us about the mind (so far)? *Cortex*, 42: 387-392, 2006.
- JACK A, SYLVESTER C and CORBETTA M. Losing our brainless minds: How neuroimaging informs cognition. *Cortex*, 42: 418-421, 2006.
- JONIDES J, NEE DE and BERMAN MG. What has functional neuroimaging told us about the mind? So many examples, so little space. *Cortex*, 42: 414-417, 2006.
- MORTON J. Brain-based and non-brain-based models of language. In Caplan D, Lecours AR and Smith A (Eds), *Biological Perspectives on Language*. Cambridge, MA: MIT Press, 1984.
- PAGE M. What can't functional neuroimaging tell the cognitive psychologist? *Cortex*, 42: 428-443, 2006.
- 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.
- SCHUTTER D, DE HAAN E and VAN HONK J. Scaling problems in the brain-mind conundrum. *Cortex*, 42: 411-413, 2006.
- SERON X and FIAS W. How images of the brain can constrain cognitive theory: The case of numerical cognition. *Cortex*, 42: 406-410, 2006.
- SMITH EE and JONIDES J. Working memory: A view from neuroimaging. *Cognitive Psychology*, 33: 5-42, 1997.
- UMILTÀ C. Localization of cognitive functions in the brain does allow one to distinguish between psychological theories. *Cortex*, 42: 399-401, 2006.
- VALLAR G. Mind, brain, and functional neuroimaging. *Cortex*, 42: 402-405, 2006.
- VANDENBERGHE R, DUNCAN J, DUPONT P, WARD R, POLINE JB, BORMANS G, MICHIELS J, MORTELMANS L and ORBAN GA. Attention to one or two features in left or right visual field: A positron emission tomography study. *Journal of Neuroscience*, 17: 3378-3750, 1997.
- WIXTED JT and STRETCH V. In defense of the signal-detection interpretation of Remember/Know judgments. *Psychonomic Bulletin & Review*, 11: 616-641, 2004.
- WOJCIULIK E and KANWISHER N. The generality of parietal involvement in visual attention. *Neuron*, 23: 747-764, 1990.

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