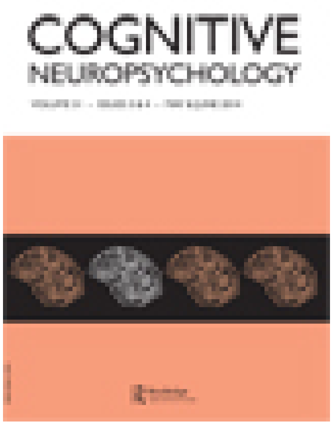

Cognitive Neuropsychology
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/pcgn20

Brain Imaging, Connectionism, and Cognitive Neuropsychology

Max Coltheart [a]
[a] Macquarie University , Sydney, Australia
Published online: 13 May 2010.

To cite this article: Max Coltheart (2004) Brain Imaging, Connectionism, and Cognitive Neuropsychology , Cognitive Neuropsychology, 21:1, 21-25, DOI: 10.1080/02643290342000159

To link to this article: http://dx.doi.org/10.1080/02643290342000159

COGNITIVE NEUROPSYCHOLOGY, 2004, 21 (1), 21–25

BRAIN IMAGING, CONNECTIONISM, AND COGNITIVE NEUROPSYCHOLOGY

Max Coltheart
Macquarie University, Sydney, Australia

I'm still an ultra-cognitive-neuropsychologist after all these years, and so my comments on Trevor Harley's most interesting book review will be from that perspective. Ultra-cognitive-neuropsychologists take cognitive neuropsychology to be a branch of cognitive psychology. The aim of cognitive psychology is to learn more about the mental information-processing systems that people use to carry out various cognitive activities. Some cognitive psychologists do that by studying the performance of people whose cognitive processing systems are normal. Others do it by studying people in whom some cognitive processing system is abnormal: Such investigators are cognitive neuropsychologists.

Cognitive neuropsychologists are not studying the brain. Investigating the brain processes upon which carrying out cognitive activities depends is a different discipline with a different name: It is called cognitive neuroscience. Some cognitive neuropsychologists are also cognitive neuroscientists because they do both kinds of work (Tim Shallice, Chris Frith, or Martha Farah, for example); some are just plain cognitive neuropsychologists (John Marshall, Alfonso Caramazza, or Max Coltheart, for example).

SHOULD THERE BE ANY "NEURO" IN COGNITIVE NEUROPSYCHOLOGY?

Harley considers this question mainly in relation to research in cognitive neuroimaging. He quotes (p. 9) an unkind epithet from Uttal (2001), who "derogatorily dubbed this emphasis on imaging 'the new phrenology'." I have come across the even less kind term "chromophrenology." Is there more to cognitive neuroimaging research than what Harley refers to as "tokenism" ("a glossy magazine might print a picture of a car because it looks good and is fun to do")? What did we learn about the mind from the Decade of the Brain?

Here we need to distinguish between two questions:

1. Has cognitive neuroscience already succeeded in using, or if not might it ever (in principle, or even in practice) successfully use, data from cognitive neuroimaging to localise cognitive processes in the brain?
2. Has cognitive neuroscience already succeeded in using, or if not might it ever (in principle, or even in practice) successfully use data from cognitive neuroimaging to make theoretical decisions entirely at the cognitive level (e.g., to adjudicate between competing information-processing models of some cognitive system)?

Fortunately, questions about what might happen with cognitive neuroimaging in the future are beyond both the scope and the space limitations of this article; so I will largely confine myself, as did Harley, to a discussion of what has happened as of the time of writing with respect to these two questions about cognitive neuroimaging.

I will consider the question of localisation first. Some of the earliest work in cognitive neuroscience had localisation of cognitive processes as an aim; for example, PET research in both St Louis (Peterson,

http://www.tandf.co.uk/journals/pp/02643294.html DOI:10.1080/02643290342000159

Fox, Posner, & Mintun, 1989; Peterson, Fox, Posner, Mintun, & Raichle, 1988) and at the Hammersmith Hospital in London (Howard et al., 1992) was aimed at localising three components of the reading system—orthography, phonology, and semantics.

The visual word form system or orthographic lexicon was localised in prestriate visual cortex by the St Louis group and in the posterior part of the left middle temporal gyrus by the London group. Phonology for reading was localised in the temporal-parietal junction by the St Louis group and in the middle part of the left superior and middle temporal gyri by the London group. Semantics was localised in prefrontal cortex by the St Louis group and in the temporal lobe by the London group. So there was simply no agreement between the two groups as to the cerebral localisation of any of these components of the reading system. Even today there is no agreement as to the localisation of components of the reading system on the basis of cognitive neuroimaging research (see, e.g., Dehaene, Le Clec, Poline, Le Bihan, & Cohen, 2002; Leff, Crewes, Plant, Scott, Kennard, & Wise, 2001).

This problem is not confined to reading. I don't know of any examples in which there is current consensus as to the cerebral localisation of any module of any cognitive system on the basis of cognitive neuroimaging data. Consider face processing, for example. Much imaging work has implicated the FFA (Fusiform Face Area; see, e.g., Kanwisher, Tong, & Nakayama, 1998) in the processing of faces. But which module of the face processing system is localised to the FFA is still not known: "What clues do the present data provide about the exact operations that are carried out in the FFA? . . . the FFA may be involved in face *detection* but not face *recognition* . . . However, the present data cannot rule out the possibility that the FFA is involved in face recognition" (Kanwisher et al., 1998, p. B9).

Perhaps I am being too impatient here; such consensus may be achieved in the future. But perhaps we should also take seriously the arguments of Van Orden and Paap (1997). It is common in contemporary cognitive modelling to view processing as cascaded (rather than thresholded) and as interactive (rather than purely feedforward). If a cognitive system does have these properties, that poses grave difficulties for the use of imaging to discover the cerebral localisation of cognitive modules. Suppose one believes that some cognitive system includes a sequence of three modules A to B to C, and one wants to localise one of these, say A, by imaging the brains of people as they carry out some task that requires module A. Because of the cascaded nature of the system, modules B and C will be activated if module A is, even if modules B and C are irrelevant as far as the task is concerned. And because of the interactive nature of the system, some of the activity in module A will be due to feedback from modules B and C. So it will not be the case that as the task is being performed all activity in the brain detected by imaging will reflect just the operation of module A in isolation. If so, how could one ever determine which parts of the brain activity here are specifically associated with module A?

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. This is why the ultra-cognitive-neuropsychologist's answer to the question "Should there be any 'neuro' in cognitive neuropsychology?" is "Certainly not; what would be the point?".

Is this extreme position tenable? If there are already any examples from the literature of cognitive neuroimaging in which some otherwise plausible information-processing model of cognition has been refuted on the basis of imaging data, then this plank of the ultra-cognitive-neuropsychological platform would have to be abandoned forthwith. Harley could not think of any such examples; nor can I[1].

As Harley says, "Clearly the identification of distinct neural structures in the brain in itself can tell us nothing about cognition. Studying the brain can only make a contribution to understanding the mind when we can relate the two in some way" (p. 8). Let's be clear what's meant by "relate the two" here. In order to localise the modules of a cognitive system, we must first know what the system's modules actually are. So we must begin with a model and then seek to do localisation research. Imaging is not *contributing* to modelling here; instead, imaging *depends* on (prior) modelling.

COMPUTATIONAL MODELLING AND COGNITIVE NEUROPSYCHOLOGY

In Harley's discussion of the role of computational modelling in cognitive neuropsychology (he claims that such modelling has been unjustly neglected), he treats as synonymous the terms "connectionism" and "computational modelling." I suggest that this is injudicious because it overlooks a crucial distinction. As he says (p. 12), in connectionist modelling "complex processes are related to low-level, neural-like processes"—the connections in such models are thought of as "neuron-like," as having some potential physical realisation. But there is a kind of computational modelling in which, although the term "connection" may be used in describing the structure of the model, the modeller does not think of individual connections as neuron-like or having any potential physical realisability. Examples of such models include the WEAVER model of spoken word production (Roelofs, 1997), the Shortlist model of spoken word recognition (Norris, 1994), and the DRC model of visual word recognition and reading aloud (Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001).

Such models are intended as descriptions of what is going on at the functional, information-processing level. They are not intended to say anything whatsoever about the neural level. In contrast, modellers who explicitly describe their work as connectionist are also liable to describe the connections in their models as "neuron-like" and to describe their models as "neurally plausible." Hence I would argue that two approaches to the computational modelling of cognition are practised, the connectionist and the nonconnectionist.

Connectionist computational modellers, in addition to treating the elements of their models as brain-like, typically develop their models via a connectionist learning algorithm such as back-propagation or Hebbian learning, typically assume that representations are distributed, and typically assume that processing is parallel.

Nonconnectionist computational modellers, in addition to treating the elements of their models as descriptions of a functional information-processing architecture rather than of some physical system such as the brain, typically develop their models by considering what the relevant empirical literature suggests abut the functional architecture of a particular cognitive processing system, make no assumptions about whether representation is distributed or local (though in this tradition models with local representations are much more common), and make no assumptions about whether processing is serial or parallel; indeed, in many such models, some parts of processing are serial (e.g., grapheme-to-phoneme translation in the DRC model) and some parts are parallel (e.g., letter identification in the DRC model).

Given this distinction, consider the following: "nevertheless, in spite of . . . the clear advantages of computational modelling, discovering the functional architecture is still the most pressing concern

[1] Note that this has *nothing to do with localisation*. No matter how many dazzling successes in localising cognitive modules using imaging had been achieved, the point stands, because such successes do not represent refutations or confirmations of information-processing models of cognition.

among cognitive neuropsychologists, and connectionism is widely, and incorrectly, disregarded" (Harley, p. 7). Two important things are clear from this quotation:

1. It explicitly identifies connectionism with computational modelling; I have offered reasons for repudiating this identification
2. It also treats the search for functional architecture and connectionist modelling as two different endeavours. I believe that is true. If it is true, however, that makes the distinction between connectionist and nonconnectionist computational modelling an even sharper one because nonconnectionist computational models of the kind I refer to above are, exactly, proposals about the functional architecture of some cognitive system or other.

Is Harley correct that connectionist modellers of cognition are not seeking to discover the functional architecture of cognition? I think that this can be demonstrated. Consider the connectionist modelling of reading aloud reported by Plaut, Seidenberg, McClelland, and Patterson (1996). They developed a three-layer neural network for reading aloud, trained by backpropagation. It did an excellent job of learning to read aloud a set of about 3000 words, including many exception words, and when tested on nonwords (absent from the training set), it also performed well. After training, then, we have a single processing system that can read aloud both exception words and nonwords, refuting the suggestion sometimes made that the ability to perform both tasks implies a dual-route structure for the reading system. Plaut et al. (1996, p. 89), however, made the following observation: Success on this task might have been achieved because the result of training may have been "that the network had partitioned itself into two subnetworks: A fully componential one for regular words (and nonwords) and a much less componential one for exception words."

This raises the important distinction between network architecture and functional architecture. It is easy to find out what the network architecture of a connectionist model is, since that is fully decided by the connectionist modeller. In the case of the Plaut et al. model, the network architecture consists of three layers, with 105 input units fully connected to 100 hidden units, which are in turn fully connected to 61 output units (with reciprocal connections from output units to hidden units in the case of the attractor version of the network). In contrast, it is very difficult to discover the functional architecture of a trained network for networks of realistic size—that is, to discover how the trained network has been structured by the learning algorithm so that it can perform the task it has learned. Trained connectionist networks thus have functional architectures, but the connectionist modellers using the learning-algorithm approach very rarely seek to discover what the functional architectures of their networks are (Plaut et al., 1996, being an exception). Hence Harley is correct in asserting that, at least in practice, seeking functional architectures and doing connectionist modelling of cognition are two separate endeavours. If the aim of cognitive psychology is to discover functional architectures of cognition, and if the aim of cognitive neuropsychology is to discover functional architectures of cognition by studying people with disorders of cognition, then why would one worry about connectionism being disregarded by cognitive neuropsychology?

Disregard of *all kinds* of computational modelling by cognitive neuropsychologists would, on the other hand, be worrying, because, as Harley says, this approach has many virtues. But I don't think there is such disregard. Computational modelling (often of a nonconnectionist kind that is specifically concerned with functional architecture) has been applied to the simulation of a number of different kinds of cognitive impairments—for example, acquired dyslexia (e.g., Coltheart, Langdon, & Haller, 1996; Coltheart et al., 2001), aphasic speech production (e.g., Foygel & Dell, 2000), and prosopagnosia (Young & Burton, 1999).

REFERENCES

Coltheart, M., Langdon, R., & Haller, M. (1996). Computational cognitive neuropsychology and acquired dyslexia. In B. Dodd, L. Worrall, &

R. Campbell (Eds.), *Evaluating theories of language: Evidence from disordered communication*. London: Whurr Publishers.

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

Dehaene, S., Le Clec, H. G., Poline, J. B., Le Bihan, D., & Cohen, L. (2002). The visual word form area: A prelexical representation of visual words in the fusiform gyrus. *NeuroReport*, *13*, 321–325.

Foygel, D., & Dell, G. S. (2000). Models of impaired lexical access in speech production. *Journal of Memory and Language*, *43*, 182–216.

Howard, D., Patterson, K., Wise, R., Brown, W., Friston, K., Weiller, C., & Frackowiak, R. (1992). The cortical localization of the lexicons. *Brain*, *115*, 1769–1782.

Kanwisher, N., Tong, F., & Nakayama, K. (1998). The effect of face inversion on the human fusiform face area. *Cognition*, *68*, B1–B11.

Leff, A. P., Crewes, H., Plant, G. T., Scott, S. K., Kennard, C., & Wise, R. J. (2001). The functional anatomy of single-word reading in patients with hemianopic and pure alexia. *Brain*, *12*, 510–521.

Norris, D. (1994). Shortlist: A connectionist model of continuous speech recognition. *Cognition*, *52*, 189–234.

Peterson, S. E., Fox, P. T., Posner, M. I., & Mintun, M. (1989). Positron emission tomographic studies of the processing of single words. *Journal of Cognitive Neuroscience*, *1*, 153–170.

Peterson, S. E., Fox, P. T., Posner, M. I., Mintun, M., & Raichle, M. E. (1988). Positron emission tomographic studies of the cortical anatomy of single-word processing. *Nature*, *331*, 585–589.

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

Roelofs, A. (1997). The WEAVER model of word-form encoding in speech production. *Cognition*, *64*, 249–284.

Uttal, W. R. (2001). *The new phrenology: The limits of localizing cognitive processes*. Cambridge, MA: MIT Press.

Van Orden, G. C., & Paap, K. R. (1997). Functional neuroimages fail to discover pieces of mind in parts of the brain. *Philosophy of Science Proceedings*, *64*, S85–S94.

Young, A. W., & Burton, A. M. (1999). Simulating face recognition: Implications for modeling cognition. *Cognitive Neuropsychology*, *16*, 1–48.

Max Coltheart
Macquarie University, Sydney, Australia
Email: max@maccs.mq.edu.au