

Chapter 14

Is there anything special about working memory?

Bradley R. Buchsbaum and Mark D'Esposito

Introduction

In a recent article entitled 'What has functional neuroimaging told us about the mind (so far)?' Coltheart (2006) concludes that the answer to this question is: 'nothing'. The essential reason for this gloomy assessment about the value of functional neuroimaging for cognitive psychology boils down to this essential argument, namely, that since psychological theories do not make predictions about the brain, it follows that such theories are necessarily consistent with all possible brain imaging results. Of course, while this statement is true, it is also tautological – i.e. it is true by definition. Coltheart's conclusion rests on the premise that psychological theories do not and cannot make predictions about the brain. A hypothetical philosopher of metaphysics might ask the question: 'What has physics told us about metaphysics?' to which he might answer that because metaphysics is the science of the non-physical, physics by definition has nothing to say about metaphysics. Unlike metaphysics and physics, however, most would agree that the study of the mind and the study of the brain are fundamentally related if not, indeed, one and the same endeavour. There is therefore absolutely no reason why psychological theories should not refer to and make explicit predictions about brain function, nor is there any reason to think such theories would, upon making contact with neuroscience, somehow cease to be 'psychological'. Indeed, the four chapters in this section all stand out as examples, contra to Coltheart claims, that psychological theories can and do make predictions about the brain, and that, in the words of Postle (Chapter 8, p. 2), 'the distinction between strictly cognitive vs. strictly neural science is often no longer appropriate'.

New insights on short- and long-term memory

One of the classic divisions used by memory researchers, as well as clinicians caring for patients with memory disorders, is that between short- and long-term memory. Short-term memory is defined as the ability to store information temporarily (for seconds) before it is consolidated into long-term memory. Clinicians often examine short-term memory with a test such as digit span (e.g. repeat these digits immediately back to me: 4, 3, 7, 1, 5, 0, 6). The concept of short-term memory has evolved into 'working memory' which refers to the temporary maintenance of information that was just experienced or

just retrieved from long-term memory but no longer exists in the external environment. These internal representations are short-lived, but can be maintained for longer periods of time through active rehearsal strategies, and can be subjected to various operations that manipulate the information in such a way that makes it useful for goal-directed behaviour. In contrast, long-term memory is defined as the ability to learn new information and recall (or recognize) this information after some time has passed. Clinicians often examine long-term memory by asking the patient to learn items that must be retrieved after an interval with some distraction (e.g. recall of three items – cat, apple, table – after one minute of performing some other task). The term amnesic syndrome, as used by clinicians, refers to the loss of long-term memory only.

The report of patient H.M. by Scoville and Milner (Scoville & Milner, 1957) provided support for this sensible distinction of memory function. Patient H.M. had intractable epilepsy and underwent bilateral surgical excision of the hippocampus and amygdala. Following surgery, H.M. suffered a dense and isolated impairment in episodic memory that still persists today. However, H.M. did not exhibit a short-term memory deficit since he was reported as having a normal digit span. In contrast, a number of years later, several patients were reported (Vallar & Baddeley, 1984; Warrington & Shallice, 1969) with the opposite deficit – impaired short-term memory and intact long-term memory. The combination of these two patterns of memory deficits – which together constituted a *double dissociation* – amounted to a powerful argument for the independence of short- and long-term memory stores in the brain. These conclusions drawn from human lesion data were further bolstered by contemporaneous information processing models in cognitive psychology that also postulated separate and independent short- and long-term memory systems (Atkinson & Shiffrin, 1968). This convergence of evidence supporting a two-system view of memory made it the dominant paradigm in memory research over the past 50 years.

The data presented in all of the chapters in this section in one way or another buck this trend in that they emphasize *commonalities* in the processes underlying short- and long-term memory rather than differences. Moreover, much of the impetus for this more unified treatment of human memory comes directly from functional neuroimaging data. Thus, Ranganath (Chapter 9) reviews the recent history of neuroimaging studies that seemed to demonstrate what had previously been nearly unthinkable – that the hippocampus is important for active maintenance of information in working memory. Ranganath further describes how these neuroimaging findings led to a renewed interest in testing patients with hippocampal lesions on tests of short-term memory, and consequently, to a reappraisal of earlier conclusions. Also, in an elegant set of functional MRI (fMRI) studies Ranganath challenged the long-held view that rehearsal in working memory plays no role in long-term memory formation. Specifically, these studies demonstrated that processing during the initial stage of working memory maintenance directly and disproportionately contributes to successful long-term memory formation. In our opinion, this neuroimaging study has told us something about the ‘mind’ by testing predictions derived from purely behavioural studies and providing data that helps to shape and modify cognitive theory.

In the chapter by Johnson and Johnson, a similar effort is made to escape the strait-jacket of the two-system view of memory. In their presentation of the Multiple-Entry, Modular (MEM) model, the authors develop a framework that consists of a set of core cognitive processes that are marshaled in various contexts and combinations in the service of cognitive control. This means that less emphasis is placed on the traditional short- and long-term memory systems, but rather on the constituent operations that are required for ‘reflective’ cognitive operations. For instance, ‘noting’, ‘refreshing’, ‘reactivation’, ‘rehearsal’, and ‘retrieval’ comprise the cognitive vocabulary of the MEM model, and each of these operations may play a role in working memory and long-term memory, depending on the context. In this chapter, they have primarily focused on one of these operations – that of ‘refreshing’ just-presented perceptual information – and have linked this operation to the dorsolateral prefrontal cortex (DLPFC). An important aspect of their work is that while the MEM model has a distinctive cognitive psychological ‘flavour’ to it, the authors have made clear predictions about its relation to functional neuroanatomy.

Indeed, the explicit neuroanatomical link drawn between ‘refreshing’ and the DLPFC in the MEM model brings it in to direct competition with the ideas of Owen and Hampshire (Chapter 7), who ascribe a similar, but more general, function to the mid ventrolateral prefrontal cortex (VLPFC). Thus, Owen and Hampshire review recent evidence that suggests that the mid VLPFC plays a specific role in *intended action* – that is, ‘any behaviour that is consciously willed by the agent responsible for carrying out that behaviour’ (p. 3). Owen and Hampshire, like the other chapters in this section, pursue a hypothesis about frontal lobe function that goes beyond the memory systems approach. For instance, they note that both short-term memory (e.g. ‘digit span’) and long-term memory tasks, as well as ‘cognitive control’ type tasks with no clear memory component at all, appear to recruit the mid VLPFC in equal measure. They also show that the mid VLPFC is active during both memory *encoding* and memory *retrieval* – two aspects of long-term memory that are often considered apart – provided that the subjects had been instructed to attend to the mnemonic stimuli with the *intention* of storing it for later use or remembering its prior occurrence. Finally, Owen and Hampshire show that in a task requiring shifts of attention both within (‘intradimensional shift’) and across (‘extradimensional shift’) stimulus categories, the mid VLPFC was most active in the latter case, wherein a complete reconfiguration of attentional set is required. This kind of cross-category set shifting, they argue, is mediated by a top-down biasing or ‘tuning’ process that modulates activity in modality-specific posterior regions, according to the currently relevant behavioural goal. It remains to be seen how other models of VLPFC function (Badre & Wagner, 2007; Thompson-Schill, D’Esposito, Aguirre, & Farah, 1997), which emphasize selection and controlled retrieval processes, can be reconciled with this view. In addition, one wonders whether the ‘refresh’ operation of the MEM model might also depend on this kind of attentional biasing; and, if so, to what extent this mechanism is a general capability of the prefrontal cortex or whether there are true dissociations between DLPFC and VLPFC as is implied by both Owen and Hampshire, and Johnson and Johnson.

In the chapter by Postle, we are shown how the study of short-term memory has benefited from an intra-disciplinary approach – from cognitive psychology, to monkey electrophysiology, to clinical neurology, to human functional neuroimaging. The Working Memory model of Baddeley and Hitch (1974) initially distinguished between short-term retention of visual and verbal information. Neurobiological investigations of the functional anatomy of the visual system, however, supported a dissociation between object and spatial processing ('what/where') streams in both perception (Ungerleider & Mishkin, 1982) and delayed response tasks (Goldman-Rakic, 1987). Postle and colleagues hypothesized that Parkinson's disease (PD) offered a good model to study this object/spatial dichotomy in the context of working memory because of its systematic pattern of neurodegeneration and because previous studies had found impairment in PD patients in tasks of spatial working memory. They found that PD patients were selectively impaired in memory for spatial location information (relative to information about object identity), arguing for a dissociation between object and spatial working memory that was dependent on the prefrontal-basal ganglia circuit that is compromised in PD. These results stimulated fMRI studies on object and spatial working memory that showed a greater role for the caudate nucleus (a region compromised in PD) for spatial working memory, as well as dissociations in the posterior neocortex. These and other studies ultimately led Baddeley and Logie to fractionate the visual subsystem of the Working Memory model into a 'visual cache' and 'inner scribe' for the representation of object and spatial information, respectively. Ultimately, then, the chapter by Postle shows how the willingness to pursue an idea using different research styles and methodologies – even those that cut across traditional disciplines – may generate new evidence and insights that can then 'ripple back' and change the very theoretical framework from which the original research question was formed.

Neuroimaging begets neuroimaging

An important thread running through each of the chapters in this section is that the models of brain function that they advance have emerged not just from ideas borrowed from cognitive psychology and basic neuroscience, but rather have (at least partly) derived from an empirical study of prior neuroimaging research. Thus, while the first generation of functional neuroimaging studies was overwhelmingly focused on 'brain mapping' and relied heavily on other disciplines for theoretical guidance, such research nevertheless supplied a mountain of empirical data from which a second generation of functional neuroimaging research could mine for patterns and commonalities that would lead to the formulation new hypotheses about the function of the brain; and thus, in the last few years, functional neuroimaging community has truly begun to 'eat its own dog food'. Indeed, three of the four articles (Ranganath, Johnson & Johnson and Owen & Hampshire) cite functional neuroimaging meta-analyses that they or others carried out as leading to new insights or contributing to the generation of new hypotheses about the role of certain brain regions in the domain of memory and cognitive control.

For instance, in trying to address the role of the DLPFC in long-term memory encoding, Ranganath and Blumenfeld examined a large number of neuroimaging studies that

had examined so-called ‘subsequent memory’ effects. What they found was that the overwhelming majority of such studies found these effects in the VLPFC while very few were found in the DLPFC. Far from concluding that the DLPFC is unimportant for long-term memory formation, the result of the meta-analysis, coupled with previous working implicated the DLPFC in relational memory processing, led Ranganath to hypothesize that the reason previous studies had not observed subsequent memory effects in DLPFC was because they had not used a task that had required relational processing at encoding. This speculation turned out to be correct, as Ranganath demonstrated that when relational or associative processing was required at encoding, activation in DLPFC predicted subsequent memory performance.

Despite the apparent success of using neuroimaging data to provide valuable insight regarding the brain basis of memory function, this method cannot provide all the answers. Other neuroscientific methods can offer different temporal and spatial resolution as well as provide data that can support different types of inferences that can be drawn from it. Undoubtedly, data obtained addressing a single question but derived from multiple methods will provide more comprehensive and inferentially sound conclusions. Functional neuroimaging studies support inferences about the association of a particular brain system with a cognitive process. However, it is difficult to prove in such a study that the observed activity is necessary for an isolated cognitive process because perfect control over a subject’s cognitive processes during a functional neuroimaging experiment is never possible. Even if the task a subject performs is well designed, it is difficult to demonstrate conclusively that he or she is differentially engaging a single, identified cognitive process. The subject may engage in unwanted cognitive processes that either have no overt, measurable effects or are perfectly confounded with the process of interest. Consequently, the neural activity measured by functional neuroimaging may result from some confounding neural computation that is itself not necessary for executing the cognitive process seemingly under study. It is important to note that the limitations in the inferences that can be drawn from functional neuroimaging studies such as fMRI apply to all methods of physiological measurement (e.g. EEG or MEG).

The inference that a particular brain region is necessary for a cognitive process cannot be made without showing that inactivating that brain region disrupts the cognitive process in question. When the results from lesion and functional neuroimaging studies are combined, a stronger level of inference emerges. If a lesion of a specific brain region causes impairment of a given cognitive process and when engaged by an intact individual, that cognitive process evokes neural activity in the same brain region, the inference that this brain region is computationally necessary for the cognitive process is stronger than data derived from each study performed in isolation. Thus, lesion and functional neuroimaging studies are complementary, each providing inferential support that the other lacks.

Although none of the chapters in this section presented data obtained from the same paradigm in both healthy individuals and those with focal lesions, several authors consider their fMRI data in the context of lesion data collected in other contexts. For example, Ranganath notes that his fMRI findings showing hippocampal activity during the

temporary retention of relational information is consistent with studies of amnesics that have found that damage limited to the hippocampus exhibit impairments in immediate memory for stimuli that require relations processing such as the locations of objects within a complex scenes (Hannula, Tranel, & Cohen, 2006). Other functional neuroimaging data presented in these chapters have not been linked to lesion data but would benefit from such a comparison. For example, two authors attribute seemingly different cognitive processes to DLPFC function. Ranganath proposes that DLPFC is involved in ‘processing of relationships between items in working memory’ whereas Johnson and colleagues proposes that the DLPFC is involved in refreshing: the act of thinking of, or foregrounding, a representation of a thought or percept which was activated just a moment earlier and has not yet become inactive. Perhaps studies of patients with focal lesions to subregions of the PFC can reconcile these different viewpoints derived from imaging data.

Conclusions

Each of the four chapters in this section illustrate how psychological theory and functional neuroimaging (and other methods as well) are no longer ‘strange bedfellows’. From the standpoint of memory research, the data presented herein is a part of a broader trend in the cognitive neuroscience community that is moving beyond the strict two-system view of memory. From the multiple encoding framework of Postle, to Ranganath’s investigations of role of the hippocampus in working memory, to Owen and Hampshire’s examination of the VLPFC in digit span and long-term encoding and retrieval, and Johnson and Johnson’s investigations of the refreshing, ‘memory’ is increasingly being seen as a by-product of some more basic neural mechanisms. Indeed, this view was advocated by Joaquin Fuster in his monograph entitled, *Memory in the Cerebral Cortex* (Fuster, 1995) which he begins with the following sentence: ‘This book is about the *memory of systems*, not about *systems of memory*’. He went on to elaborate with a paragraph that captures the spirit of the four chapters in this section and with which we will close: ‘Memory is a functional property, among others, of each and all of the areas of the cerebral cortex, and thus all cortical systems. This cardinal cognitive function is inherent in the fabric of the entire cortex and cannot be ascribed exclusively to any of its parts. Furthermore, as the cortex engages in representing and acting on the world, memory in one form or another is an integral part of all operations. In this respect, what distinguishes one cortical area from another is its *kind* (i.e. the content and history) of memory; in the temporal domain, it is the *state* of memory, active or inactive or somewhere in between. Thus, as one of the cognitive functions of the cerebral cortex, memory is global and nonlocalizable. Its most concrete contents, however, those that are inextricable from specific sensory or motor functions, are well localized.’

References

- Atkinson, R. C. & Shiffrin, R. M. (1968). Human Memory: A proposed system and its control processes. In K. W. Spence (Ed.), *The Psychology of Learning and Motivation: Advances in Research and Theory* (Vol. 2, pp. 89–195). New York: Academic Press.

- Baddeley, A. D. & Hitch, G. J. (1974). Working Memory. In G. Bower (Ed.), *The Psychology of Learning and Motivation* (Vol. 7, pp. 47–90). New York: Academic Press.
- Badre, D. & Wagner, A. D. (2007). Left ventrolateral prefrontal cortex and the cognitive control of memory. *Neuropsychologia*, *45*(13), 2883–2901.
- Coltheart, M. (2006). What has functional neuroimaging told us about the mind (so far)? *Cortex*, *42*(3), 323–331.
- Fuster, J. M. (1995). *Memory in the Cerebral Cortex*. Cambridge: The MIT Press.
- Goldman-Rakic, P. S. (1987). Circuitry of primate prefrontal cortex and regulation of behavior by representational memory. In F. Plum (Ed.), *Handbook of Physiology – the Nervous System* (Vol. 5, pp. 373–417). Bethesda, MD: American Physiological Society.
- Hannula, D. E., Tranel, D., & Cohen, N. J. (2006). The long and the short of it: relational memory impairments in amnesia, even at short lags. *Journal of Neuroscience*, *26*(32), 8352–8359.
- Scoville, W. B. & Milner, B. (1957). Loss of recent memory after bilateral hippocampal lesions. *Journal of Neurology, Neurosurgery and Psychiatry*, *20*(1), 11–21.
- Thompson-Schill, S. L., D'Esposito, M., Aguirre, G. K., & Farah, M. J. (1997). Role of left inferior prefrontal cortex in retrieval of semantic knowledge: a reevaluation. *Proceedings of the National Academy of Sciences of the United States of America*, *94*(26), 14792–14797.
- Ungerleider, L. & Mishkin, M. (1982). Two cortical visual systems. In D. Ingle, M. A. Goodale, & R. J. W. Mansfield (Eds), *Analysis of Visual Behavior* (pp. 549–586). Cambridge, MA: MIT Press.
- Vallar, G. & Baddeley, A. (1984). Fractionation of Working Memory: Neuropsychological Evidence for a Phonological Short-Term Store. *Journal of Verbal Learning and Verbal Behavior*, *23*, 151–161.
- Warrington, E. & Shallice, T. (1969). Selective Impairment of Auditory Verbal Short-Term Memory. *Brain*, *92*, 885–&.

