Bilateral Generic Working Memory Circuit Requires Left-Lateralized Addition for Verbal Processing

According to the Baddeley–Hitch model, phonological and visuospatial representations are separable components of working memory (WM) linked by a central executive. The traditional view that the separation reflects the relative contribution of the 2 hemispheres (verbal WM—left; spatial WM—right) has been challenged by the position that a common bilateral frontoparietal network subserves both domains. Here, we test the hypothesis that there is a generic WM circuit that recruits additional specialized regions for verbal and spatial processing. We designed a functional magnetic resonance imaging paradigm to elicit activation in the WM circuit for verbal and spatial information using identical stimuli and applied this in 33 healthy controls. We detected left-lateralized quantitative differences in the left frontal and temporal lobe for verbal > spatial WM but no areas of activation for spatial > verbal WM. We speculate that spatial WM is analogous to a “generic” bilateral frontoparietal WM circuit we inherited from our great ape ancestors that evolved, by recruitment of additional left-lateralized frontal and temporal regions, to accommodate language.

Keywords: functional MRI, language, laterality, neuroimaging, phonological loop, visuospatial sketchpad, working memory

Introduction

Inputs to working memory (WM) arise either from current perception or from activation of long-term memory (LTM) and are retained temporarily in a highly accessible state. Thus, WM is an interface between perception, LTM, and action (Baddeley 2003) and underlies human thought processes by supporting maintenance and manipulation of information in a labile store.

According to the multicomponent model of Baddeley and Hitch (1974), WM comprises an attention-controlling mechanism, the central executive and 2 subsidiary slave systems, the phonological loop (verbal WM), and the visuospatial sketchpad (object and spatial WM). The visuospatial sketchpad is involved in mental rotation, scanning, and comparison (Logie 1995). In contrast to verbal WM, which is involved in language acquisition (Baddeley et al. 1998), spatial and object WM is well documented in nonhuman primates (Davachi et al. 2001; Castner et al. 2004; Inoue et al. 2004). Due to differential WM abilities in human and nonhuman primates, the distinction between neural correlates of verbal and spatial WM has become a prime candidate for understanding WM-dependent language acquisition.

Early studies appeared to separate verbal WM in the left from spatial WM in the right hemisphere (Jonides et al. 1993; Paulesu et al. 1993). Based on these early studies, a case was made for a qualitative domain (verbal/spatial)—dependent dissociation of hemispheric activation. However, later studies, involving information maintained as well as manipulated in WM, found bilateral activation for both verbal and spatial WM tasks (D’Esposito et al. 1998; Smith and Jonides 1998; Smith and Jonides 1999; Cabeza and Nyberg 2000; D’Esposito et al. 2000). Although the left–right dissociation remains obscure, it has become increasingly clear that both the prefrontal and the parietal regions are important parts of the WM circuit. Other parts of the network include the cingulate, occipital, and cerebellar regions (Cabeza and Nyberg 2000).

However, few imaging studies have examined both verbal and spatial WM tasks on the same group of subjects. Those that did used physically different stimuli for verbal and spatial tasks, for example, alphabets for verbal and blobs for spatial (D’Esposito et al. 1998). Differences in such studies may be partially attributable to perceptual stimulus rather than domain. To circumvent this problem, Smith et al. (1996) used identical stimuli (letters) for verbal and spatial task. Although hemispheric dissociation was reported, this was by visual comparison of activation maps rather than quantitative assessment.

Only 2 studies, Nystrom et al. (2000) and Walter et al. (2003), used within-subjects design, with physically identical stimuli and direct statistics, to differentiate between verbal and spatial WM. Nystrom found 5 regions of activation spanning the frontoparietal region bilaterally in both verbal and spatial tasks. The mean functional magnetic resonance imaging (fMRI) signal was higher in the spatial task for all 5 regions, but there was no evidence for domain-dependent hemispheric specialization, neither absolute (dissociation) nor relative (dominance). A possible explanation is that volunteers used a verbal strategy to remember locations. Nystrom used evenly spaced 18 positions instead of 8 (compass) or 12 (clock) to prevent this, but it is still possible that volunteers used the “nearest hour” clock position to phonologically code for the spatial cues. In addition, the sample size of 7 was small. Walter et al. (2003) conducted 2 fMRI experiments with larger numbers (13 and 15) and spatial locations organized pseudorandomly. Walter et al. also found a bilateral frontoparietal network for both conditions but with evidence for domain dominance, verbal WM being supported preferentially by left ventral prefrontal cortex compared with right dorsal prefrontal cortex preferentially supporting spatial WM. Unlike domain dissociation (complete WM type-dependent hemispheric split) of early WM maintenance studies, the finding of Walter et al. was quantitative, that is, relative specialization of the hemispheres for verbal/spatial domain.

The review of Cabeza and Nyberg (2000) shows overlapping neural circuits for verbal and spatial WM. Brodmann’s Area (BA) 6 is active bilaterally for both verbal and spatial WM. The same can be said for BA 9 and 46 but only for tasks involving manipulation. Of the prefrontal areas, only BA 44 seems to be verbal WM specific. For the parietal areas (BA 7 and 40), Cabeza and Nyberg
Methods

stimuli for both verbal and spatial WM. However, on thorough inspection of the verbal WM studies included, specially those involving manipulation, most show bilateral activation, some left-lateralized, and a couple even right-lateralized activation of BA 7 and BA 40 (Cabeza and Nyberg 2000, Table 5).

Based on these findings of largely overlapping bilateral activation for both verbal and spatial WM, along with some evidence of relative hemispheric specialization, we speculate that a generic WM network possibly preceded (in an evolutionary sense) the relatively specialized verbal and spatial WM circuits in Homo sapiens. This generic or prototype network could have been inherited from our great ape precursors, followed by structural and functional reorganization due to evolutionary pressures. However, the primary role of verbal WM is language acquisition (Baddeley et al. 1998), a human-specific characteristic (Bickerton 1990). Thus, any generic WM circuit inherited from lower primates should be more similar to spatial WM than verbal. This will also be in keeping with the fact that unlike verbal WM, spatial WM is well documented in nonhuman primates (Davachi et al. 2001; Castner et al. 2004; Inoue et al. 2004). To accommodate language, the generic WM must have evolved. So if spatial WM is akin to a prototype and verbal WM is a modified version of it, both verbal and spatial WM circuits should share overlapping frontotoparietal circuits. The prototype spatial WM may be hypothesized to have undergone specialization to allow acquisition of the phonological component of language, predictably as a left-lateralized adjunct system for maintaining and manipulating information in verbal WM. The specialization of the shared bilateral frontotoparietal circuit (prototype WM) should be apparent in a neuroimaging study, as left-sided domain dominance for verbal WM. To test the hypothesis, we carried out a within-subjects fMRI experiment on 33 healthy volunteers, using variants of the “n-back” task with physically identical stimuli for both verbal and spatial domains.

Subjects

Informed consent was obtained from 33 neurologically intact individuals (16 male and 17 females) in the age range of 20–50 years (mean age of 36 years). All participants were right handed on the Annett (1970) handedness questionnaire (scores of 20–24). Exclusion criterion included a lifetime Diagnostic and Statistical Manual of Mental Disorders (DSM IV) psychiatric diagnosis and neurological illness. All volunteers were screened with an abbreviated version of the Structured Clinical Interview from the DSM IV and the standard magnetic resonance imaging (MRI) safety questionnaire. Subjects were within the age range of 20–50 and had completed a minimum of 12 years of education, and all of them spoke fluent English. The project was approved by Oxfordshire Clinical Research Ethics Committee (Ethics No. C02.086).

Paradigms

Variants of the n-back task were used to map out verbal and spatial WM. n-back tasks involve continuous monitoring of a series of stimuli. Volunteers respond whenever a stimulus is presented that is the same as the one presented “n” trials previously. As shown in Figure 1, n is a prespecified integer varying between 0 and 3, increasing parametrically in levels of difficulty (Morris and Jones 1990). The 2- and 0-back variants of n back were used to elicit the effect of WM load in the verbal and spatial domains.

In all the 4 tasks, volunteers were exposed to English alphabets (26 letters) appearing in different positions on a 5-by-5 matrix (25 positions). Presentation and physical parameters of the verbal and spatial tasks were identical. Because letters were used to designate positions in the spatial WM task, it ensured that activation differences were not confounded by differences in perceptual stimuli. The 0-back task only involved monitoring. Subjects were instructed to push a button every time they saw a predefined stimulus. In the case of verbal 0 back, it was letter V and for spatial, it was any letter in the center square. Other than the WM load, the 0-back task was similar in every aspect to the 2-back task. In the 2-back task, subjects were required to push a button every time they saw a stimulus same as the one 2 stimuli back (not the previous stimuli but the one before that). For verbal 2 back, it was a comparison between letters, irrespective of the position in which they were displayed, and for spatial, the comparison was between the positions at which letters were displayed on the grid, irrespective of which letter it was. This meant that at any single point in time, a subject would hold in his WM 3 stimuli, 2 immediately preceding and the new item for comparison. Accurate task performance will demand continuous updating of the 3 items in the WM store. In order to prevent subjects from bypassing this computational aspect of the 2-back task by using strategies like recency, similar variables were added in 1-back and 3-back position as foils.

One major concern was to prevent use of phonological codes in the spatial WM task. It is well established that subjects tend to remember positions according to the hours of a clock or positions of a compass. This led us to consider a 5-by-5 matrix as it is almost impossible for subjects to phonologically code the 2nd and 4th rows and columns due to lack of descriptive names. We cross-checked this during pilot studies, and all subjects reported that they were focusing their attention on the area in which they were expecting a letter to be displayed (a spatial strategy). Another concern was the use of a visual strategy (letter shape comparison) in the verbal WM task. Although all subjects reported subvocal rehearsal in pilot studies as a cross-check, we placed similarly shaped letters (B, D, and P) as foils in 2-back fashion. To further explore variance in strategies used to perform the tasks, we debriefed subjects regarding particular strategies used in each task after the scan using a 5-item questionnaire. The debriefing also covered perceived comparative difficulty, effort needed, and their subjective opinion on performance during the tasks.

All subjects received training on the WM tasks. The behavioral data from the practice sessions were examined for a minimum of 80% accuracy. This was set as an inclusion criterion as it ensured continuous task engagement as well as exclusion of volunteers using strategies like recency, phonological codes in spatial task, and letter shape comparison in verbal task.

Inside the scanner, the tasks were presented in 2 sessions, verbal and spatial. To eliminate order effects, the sequence of the sessions was block randomized using a block size of 4. Each session comprised 6 blocks, 3 of 2 back and 3 of 0 back alternating as shown in Figure 2. Each block comprised 20 stimuli and lasted 1 min (500-ms stimulus display: 2500-ms interstimulus delay). In each 2-back block, there were 3 foils for similar shape and 4 recency foils (2 in 1-back and another 2 in 3-back position). A rest period of 15 s in between each block allowed the blood flow to normalize. Once inside the scanner, volunteers practiced the active task for 1 min before starting image acquisition. This provided a chance to refresh the instructions.

fMRI Data Acquisition

Imaging was performed on a 1.5 T Siemens Sonata Scanner at University of Oxford Centre for Clinical Magnetic Resonance Research. The scan
sequence consisted of echo planar imaging (EPI) sequences for each of the functional tasks, fast low angle shot sequence for a high-resolution structural scan, and a gradient recalled echo field mapping to correct for magnetic field inhomogeneity. The functional scan parameters were as follows: voxel size = 3 mm isotropic, time repetition = 3000 ms, time echo = 50 ms, flip angle = 90 degrees. In each EPI sequence, the scanner collected 150 whole brain volumes.

Analysis

Behavioral Data
The percentage accuracy and the mean response time (RT) was calculated for each individual on each task and then combined into group averages. Analysis of variance (ANOVA) was used to reveal significant variance in accuracy and RTs in the verbal and spatial tasks (both 2 and 0 back).

fMRI Data
fMRI data were analyzed using FMRI Expert Analysis Tool (FEAT Version 5.1), part of the FMRIB Software Library. Prestatistics involved brain extraction (Smith 2002), motion correction using Motion Correction FMRIB's (functional magnetic resonance imaging of the brain) Linear Registration Tool with rigid body transformations (Jenkinson and Smith 2001; Jenkinson et al. 2002), spatial smoothing to increase the signal-noise ratio (Gaussian kernel of full width half maximum of 5 mm), grand mean scaling, and high-pass temporal filtering (Gaussian-weighted least square fit straight line fitting, with sigma = 112.5 s) to remove low-frequency drifts.

Statistical Analysis
Our analysis conformed to a mixed-effects analysis of task-related effects over subjects. This proceeded in 2 stages. First, we computed the parameters of a general linear convolution model for each subject. Contrasts of parameter estimates (PEs) summarizing the task-specific memory effects (verbal and spatial tasks) were computed and entered into t-tests at the second (between subject) level using a mixed-effects model. This allowed us to generalize our inferences about domain-specific effects to the population from which our subjects came. We tested for the simple main effects of verbal and spatial WM using 1-sample t-tests and for the interaction between domain and memory by comparing the verbal and spatial effects with a 2-sample t-test. Z (Gaussianized T/F) statistic images were thresholded using clusters determined by Z > 2.3 and a (corrected) cluster significance threshold of P = 0.05 (Worsley 2001). Registration to standard space was carried out using FMRIB’s Linear Image Registration Tool (Jenkinson and Smith 2001; Jenkinson et al. 2002).

Results

Behavioral Data
Two of the 33 volunteers were excluded from the final analysis of both behavioral and MR data as inside the scanner they got the task sequence wrong (performed the 0 back in place of the 2 back and vice versa resulting in very low accuracy scores). The group accuracy means and RTs with 95% confidence intervals in the 4 tasks for the 31 volunteers (15 males, 16 females) are shown in Table 1.

As expected, there was a significant effect of WM load on both accuracy ($F_{1,30} = 43.012; P < 0.0001$) and RT ($F_{1,30} = 36.151; P < 0.0001$). The difference between verbal and spatial tasks was not significant for accuracy ($F_{1,30} = 1.716; P = 0.2$) but was significant for RT (verbal > spatial; $F_{1,30} = 7.483; P = 0.01$). There was no stimulus by WM load interaction seen for accuracy ($F_{1,30} = 2.339; P = 0.137$), but a significant interaction was present for RT ($F_{1,30} = 4.65; P = 0.03$).

In the verbal WM task, none of the subjects reported using letter shape as a means of remembering letters. This was corroborated by the fact that there were no excess false-positive responses (range of 0–2 out of a maximum of 9 foils) for the similarly shaped stimuli (e., B, D, P), which had been placed as foils in 2-back positions. In the spatial WM task, none of the subjects reported attempts to verbalize positions. The reason cited was the absence of descriptive names for positions in the 2nd and the 4th rows and columns. We also did not detect a higher rate of false-positive responses (range of 0–2 out of a maximum of 12 foils) to the recency foils in any subject. Although it is impossible to completely rule out the use of recency as a strategy, this finding does show that its use was negligible.

Table 1

<table>
<thead>
<tr>
<th>Performance</th>
<th>Mean accuracy in % (95% confidence interval)</th>
<th>Mean RT in ms (95% confidence interval)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Domain</td>
<td>Verbal</td>
<td>Spatial</td>
</tr>
<tr>
<td>WM load</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 back</td>
<td>91.46 (88.6–94.3)</td>
<td>93.46 (91–95.9)</td>
</tr>
<tr>
<td></td>
<td>796 (700–891)</td>
<td>582 (502–763)</td>
</tr>
<tr>
<td>0 back</td>
<td>99.56 (98.1–100)</td>
<td>99.46 (99–99.9)</td>
</tr>
<tr>
<td></td>
<td>589 (533–646)</td>
<td>558 (504–613)</td>
</tr>
</tbody>
</table>

Figure 2. Timeline of EPI sequences (A, rest—15 s; B, 2-back task—60 s; C, 0-back task—60 s).
From the paired t-test, the verbal WM > spatial WM contrast (Fig. 3) revealed statistically significant activation primarily in the left hemisphere, which included the inferior frontal gyrus (BA 44, 45, 46), middle frontal gyrus (BA 6), precentral gyrus (BA 6), medial superior frontal gyrus (BA 6, 8), superior and middle temporal gyrus (BA 22), inferior temporal gyrus (BA 21, 37), anterior cingulate (BA 32), insula, parahippocampal gyrus (BA 34), and lentiform nucleus. On the right, activation was limited to the anterior cingulate (BA 32) and parahippocampal gyrus (amygdala). However, spatial WM > verbal WM did not reveal any statistically significant area of activation. Further details (Talairach coordinates and Z statistics) on peak activation are provided in Table 2.

To further substantiate the lateralized findings in verbal WM > spatial WM, post hoc region of interest (ROI) asymmetry analysis was carried out. From the verbal > spatial PEs in Figure 3C, 3 areas were extracted (Fig. 4), the first comprising primarily the middle frontal gyrus (parts of BA 6, 9, 44, 45, 46—“frontal”), the second comprising posterior middle and inferior temporal gyrus (parts of BA 21, 22, 37—“temporal”), and the third comprising inferior frontal gyrus, insula, and superior temporal gyrus (parts of BA 22—“frontotemporal”). Reciprocal masks were created for the right hemisphere by flipping the data, resulting in 6 ROIs (Fig. 4). Mean percentage signal change in blood oxygen level-dependent activation in verbal and spatial WM tasks was calculated in each mask for every volunteer. Activation across the anterior cingulate, left cingulate, and left medial frontal gyrus was not included in this analysis due to proximity to the midline. ANOVA was computed to detect significant effects of domain (verbal–spatial), hemisphere (left–right), and their interaction. The results showed a highly significant effect of domain in all 3 ROIs, frontal ($F_{1,30} = 13.186$;
Discussion

Based on neuroimaging studies, Lieberman (2002) proposes that complex behavior like walking or language cannot be restricted to specific anatomical sites. They map onto multiple foci that are networked together. The findings of the current study are in agreement with the concept of a distributed network for complex operations. As expected (Cabeza and Nyberg 2000), both verbal and spatial domains activated a common bilateral frontoparietal circuit. The overlap of the 2 domains adds credence to the concept of a generic WM. Although there were hemispheric differences in extent and magnitude of activation (Figs 3 and 4), involvement of both hemispheres in verbal as well as spatial WM rules out domain-based hemispheric dissociation. As reflected by Z scores, activation for verbal WM exceeded that for spatial WM, the differences being statistically significant only in the left hemisphere. The absence of any areas of activation in the spatial > verbal contrast supports the idea that the spatial WM network is akin to the postulated generic WM.

The question arises that in a within-subjects experiment, showing no significant difference in accuracy in between tasks that have been extensively matched, how does a significant difference arise in RT. Subjects approximately took 100 ms longer to respond in the verbal WM task. The extra 100 ms correspond to either involvement of a larger neural circuitry for verbal WM or more time-intense processing in the generic WM circuit while performing verbal WM tasks. We propose that the additional recruitment of left-sided frontotemporal areas in the verbal WM task, as shown by the verbal > spatial contrast, is responsible for the prolonged RT in verbal WM task. These additional areas may also be interpreted as the predicted modification to the generic WM network to support the primary role of verbal WM—language acquisition.

Table 2

<table>
<thead>
<tr>
<th>Brain region (hemisphere)</th>
<th>Brodmann’s area</th>
<th>Z score</th>
<th>Talairach coordinates</th>
</tr>
</thead>
<tbody>
<tr>
<td>Superior frontal gyrus (L)</td>
<td>6, 8</td>
<td>4.64</td>
<td>-2 4 64</td>
</tr>
<tr>
<td>Middle frontal gyrus (L)</td>
<td>9</td>
<td>4.47</td>
<td>-46 14 30</td>
</tr>
<tr>
<td>Inferior frontal gyrus (L)</td>
<td>44, 45</td>
<td>4.39</td>
<td>-56 14 4</td>
</tr>
<tr>
<td>Caudate (R)</td>
<td>NA</td>
<td>4.34</td>
<td>16 -2 22</td>
</tr>
<tr>
<td>Superior temporal gyrus (L)</td>
<td>22</td>
<td>4.28</td>
<td>-54 2 -4</td>
</tr>
<tr>
<td>Precentral gyrus (L)</td>
<td>6</td>
<td>4.24</td>
<td>-56 2 38</td>
</tr>
<tr>
<td>Lentiform nucleus (L)</td>
<td>NA</td>
<td>4.06</td>
<td>-20 4 4</td>
</tr>
<tr>
<td>Anterior cingulate (R)</td>
<td>32</td>
<td>3.82</td>
<td>10 38 26</td>
</tr>
<tr>
<td>Inferior temporal gyrus (L)</td>
<td>21, 37</td>
<td>3.80</td>
<td>-64 -50 12</td>
</tr>
<tr>
<td>Inferior frontal gyrus (L)</td>
<td>45</td>
<td>3.60</td>
<td>-48 28 16</td>
</tr>
<tr>
<td>Superior temporal gyrus (L)</td>
<td>22</td>
<td>3.52</td>
<td>-64 -52 10</td>
</tr>
<tr>
<td>Parahippocampal gyrus (R)</td>
<td>Amygdala</td>
<td>3.50</td>
<td>28 -6 -14</td>
</tr>
<tr>
<td>Middle temporal gyrus (L)</td>
<td>22</td>
<td>3.36</td>
<td>-60 -28 8</td>
</tr>
<tr>
<td>Anterior cingulate (L)</td>
<td>32</td>
<td>3.19</td>
<td>-10 44 12</td>
</tr>
<tr>
<td>Parahippocampal gyrus (L)</td>
<td>34</td>
<td>3.13</td>
<td>-26 6 -20</td>
</tr>
<tr>
<td>Cingulate gyrus (L)</td>
<td>32</td>
<td>2.67</td>
<td>-10 18 30</td>
</tr>
</tbody>
</table>

NA, not available.

Figure 4. Mean percentage change in PEs in the frontal (1), temporal (2), and frontotemporal (3) regions with standard deviations in left and right hemisphere during verbal and spatial WM. Verbal WM gives rise to leftward asymmetry (statistically significant in all 3 ROIs, $P \leq 0.05$) in comparison to rightward for spatial (statistically significant only in temporal, $P \leq 0.02$).
anatomy supplements our theory of left-lateralized recruitment of resources to a generic WM network to accommodate language acquisition in humans.

Consistent with Walter et al. (2003), we found activation in left BA 44. This area covering part of Broca's area not only is essential for speech production (Broca 1861) but, along with BA 6, also has been implicated in the articulatory rehearsal process (Paulesu et al. 1993; Awh et al. 1996). Lesions in this area result in Broca's aphasia (speech is difficult to initiate, nonfluent, labored, and with prominent agrammatism, language is reduced to disjointed words and sentence construction is poor—Broca 1861). Embick et al. (2000) found direct evidence of syntactic specialization of Broca's area in an fMRI study comparing sentences with either grammatical or spelling errors. Volume of pars triangularis is reported to be leftwardly asymmetric (Foundas et al. 1998), and the difference holds true for BA 44 even when cytoarchitectonic borders are used instead of sulcal landmarks (Amunts et al. 1999). Foundas et al. (1998) also found higher cell densities in BA 44 in all 5 male brains and 3 out of 5 female brains.

Other frontal areas found to be active in the verbal > spatial contrast include left BA 45, which is also part of Broca's area, and left BA 9 and 46. The latter 2 areas are consistently active in contrast include left BA 45, which is also part of Broca's area, significantly active in both verbal and spatial tasks (2 back hemispheres. Lateral and medial parietal cortices were in attentional processes was active in both left and right hemispheres. Lateral and medial parietal cortices were significantly active in both verbal and spatial tasks (2 back > 0 back contrasts). Previous studies have proposed the left inferior parietal lobule (BA 40 and 7) as the site of the passive phonological store (Paulesu et al. 1993; Awh et al. 1996). Although Z scores in this region were more for verbal WM ($Z = 7.27$) than spatial WM ($Z = 6.57$), the difference was not statistically significant. This is in keeping with the review of Cabeza and Nyberg (2000) that showed a mix of lateralized as well as bilateral findings for both verbal and spatial WM.

In the temporal lobe, activation was seen in parts of left BA 21, 22, and 37 in the verbal > spatial contrast. Posterior part of BA 22, also known as Wernicke's area, is well recognized for language comprehension. Lesions in this area result in Wernicke's aphasia (impairment of language comprehension, speech remains fluent with normal syntax but has no recognizable meaning—Wernicke 1874). Geschwind and Levitsky (1968) found marked anatomical asymmetries between left and right temporal lobes in human, the left planum temporale (PT), on average is one-third longer. Because PT is a key site within Wernicke's area, consistent leftward asymmetry in this region substantiates the current findings. Along with BA 22, BA 21 has also shown to be involved in word recognition. However, written word recognition seems to be exclusive left lateralized in comparison to spoken word that appears to be bilateral (Cabeza and Nyberg 2000). In a study of visual word form area, Vigneau et al. (2005) found significant activation in the left posterior middle temporal gyrus for word versus nonword reading. This region is very close to the BA 37 activation in the verbal > spatial contrast in this study.

In the spatial > verbal WM contrast, our findings differ significantly from Walter et al. These authors found right inferior and middle frontal gyrus, right inferior parietal cortex, and both right and left precuneus to be active in their spatial > verbal contrast in their first experiment and right parietal cortex and right middle frontal gyrus in their second (Walter et al. 2003), whereas we found no significant areas of activation at all. This is the case in spite of using the same thresholds (clusters determined by $Z > 2.3$ and a corrected cluster significance of $P = 0.05$) on a much larger sample size of 31 in our experiment compared with samples of 13 and 15 in the 2 experiments of Walter et al.

Baddeley et al. (1998) make a strong case for the phonological loop as the language acquisition device proposing that its primary purpose is to store unfamiliar sound patterns of a brief and novel speech event (a new word), while more permanent memory traces are established. Thus, the ability to repeat a string of digits (e.g., phone number) is an incidental benefit of a more fundamental human capacity—the ability to acquire new words.

Because words form the building blocks of speech, vocabulary acquisition is central to language learning. There appears to be a strong association between measures of phonological loop efficiency and vocabulary knowledge (Baddeley et al. 1998). The direction of causality, phonological loop capacity influences acquisition of new words, was established by studies involving cross-lagged correlational analysis of longitudinal data (Gathercole and Baddeley 1989; Gathercole et al. 1992). The hypothesis held true in controlled environments (Gathercole and Baddeley 1990; Gathercole et al. 1997) for foreign (Service and Kohonen 1995) and second language learning (Cheung 1996). However, there is ample evidence to suggest that a child's existing knowledge of the structure of language does play a part as well (Gathercole 1995; Gathercole et al. 1997).

If language is human specific and verbal WM is the language acquisition device, then the latter would be absent in lower primates. A compelling case has been made against the presence of the core features of human language even in the most specialized forms of animal communication (Bickerton 1990). Language has been described as the last of the 8 major transitions in evolution (Maynard-Smith and Szathmary 1995 with the claim that it defines the species boundary of H. sapiens (Crow 2002). These arguments highlight the role of verbal WM from an evolutionary as well as cognitive perspective.

However, spatial and object WM (WM for where and what, respectively) are both documented in nonhuman primates (Davachi et al. 2001; Castner et al. 2004; Inoue et al. 2004). Rhesus monkeys retain visual information (memory for what) in the absence of spatial demands as well as spatial information (memory for where) in the absence of visual demands. Thus, the neural circuitry for visuospatial WM is present in lower primates. However, when monkeys are required to coordinate visual and spatial inputs (what was where), by contrast with humans, their performance is no better than chance (Washburn et al. 2003). According to the authors, humans use verbal WM to maintain and cross-reference visual and spatial information. Because verbal WM is intimately associated with language acquisition, one would intuitively expect the evolution of the neural circuitry for the same to predate spoken language. However, in a broader framework of language as a representational system (e.g., cave paintings), the other possibilities of coevolution of verbal WM and linguistic faculties or the latter preceding the former cannot be discounted.

Based on the above discussion, we propose that verbal WM is comparatively recent in evolution. It is possible that evolutionary innovations of the spatial WM network underlie the left-hemispheric differences in the magnitude of activation that we have detected. It would appear that adaptation of this network
depends upon a new principle—lateralization of the phonological engram to the left. The question arises of the physiological basis of left lateralization. Humans and chimpanzees share a common pre-hominid precursor (Ruvolo 1997; Deinard and Kidd 1999). However, relative to body size, the human brain is about 3 times larger than the chimpanzee’s (Gilissen 2001). But there is no corresponding increase in corpus callosum fiber size. In both humans (Aboitiz et al. 1992) and macaques (Lamantia and Rakic 1990), most myelinated callosal axons are below 1 μm in diameter. Thus, an increase in brain size without increase in callosal fiber thickness implies an increase in time taken for interhemispheric transfer. The average-sized myelinated fiber interconnecting the temporal lobes in humans would have a conduction time of over 25 ms. To maintain processing speed in the relatively larger human brain, it is imperative that groups of like function get clustered together in 1 hemisphere (Ringo et al. 1994). Other mammals, including nonhuman primates, have contralateral dominance in primary auditory cortex. However, in humans, activation is greater in left Heschl’s gyrus in comparison to right irrespective of ear of presentation of tone (Devlin et al. 2003). The authors conclude that the functional lateralization for primary auditory cortex may have contributed to the evolution of a unique role for the left hemisphere in language processing. It is interesting to note that in the verbal > spatial contrast, the 2 important language areas, Broca’s and Wernicke’s, are significantly active. We propose that the additional recruitment of these areas for verbal WM is an evolutionary modification of a generic WM circuit and interconnected by the arcuate fasciculus; they form a left-lateralized time- and energy-efficient network for language acquisition.

Conclusion
This is the largest within-subjects fMRI study comparing verbal and spatial WM to date. A bilateral frontoparietal network was found for both verbal and spatial WM. Although no domain-dependent local dissociation was detected, there was significant quantitative difference in the magnitude of activation. Verbal WM appeared to draw on more neural resources, lateralized to the left hemisphere. No areas of significantly greater activation for spatial WM in comparison to verbal WM were detected. The findings can be explained on the basis that the neural correlates of spatial WM are akin to a generic WM circuit. To accommodate language acquisition, the generic WM has undergone quantitative evolutionary changes in the left hemisphere to allow for the development of verbal WM.

Funding
Oxford Radcliffe Hospitals Charitable Funds (Fund No. 8086).

Notes
Sincere thanks to Prof. Phillip Cowen for his support and guidance and Jane Francis for conducting the MRI scans. Conflict of Interest: None declared.

Address correspondence to Dr Mannaan Kar Ray, Schizophrenia, A National Emergency, Prince of Wales International Centre for SANE Research, Warneford Hospital, Warneford Lane, Oxford OX3 7JX, UK. Email: mannaan.ray@psych.ox.ac.uk.

References


Address correspondence to Dr Mannaan Kar Ray, Schizophrenia, A National Emergency, Prince of Wales International Centre for SANE Research, Warneford Hospital, Warneford Lane, Oxford OX3 7JX, UK. Email: mannaan.ray@psych.ox.ac.uk.


