Review

Visual and spatial working memory: From boxes to networks

Hubert D. Zimmer

Department of Psychology, Brain & Cognition Unit, Saarland University, 66041 Saarbruecken, Germany

ABSTRACT

It is shown that visuo-spatial working memory is better characterized as processes operating on sensory information (visual appearance) and on spatial location (environmental coordinates) in a distributed network than as unitary slave system. Results from passive (short-term) and active memory tasks (imagery) disclose the properties (capacity, content) and the components of this network. The prefrontal cortex is a control structure (dorsal prefers active, ventral passive tasks) and it contributes to spatial memory by a prospective spatial code (eye movements). Visual appearance (including dynamic aspects) is represented as features and object files (bound features) within content-specific areas in the ventral occipital cortex. Spatial coordinates are represented in the parietal cortex (modality-unspecific), when used in spatio-temporal tasks (Corsi) they are closely related to attention. Imagery of objects activates occipito-temporal structures, spatial transformations and mental rotation the parietal cortex (specifically the intraparietal sulcus). Perception, working memory, and imagery use the same neural network. Differences between the tasks are explained by different demands and states of the neural network, and differences in the configuration of the anterior–posterior neural circuits.

© 2008 Elsevier Ltd. All rights reserved.
Part structures of the cognitive system that temporarily maintain information for further processing during perception and action are considered as working memory (WM). The term was originally introduced to highlight that this type of memory is active remembering and not simply a passive storage as it was investigated in short-term memory research (STM). In spite of this, in many experiments on WM, only temporary memory was tested without any additional active processing—except for some kinds of rehearsal mechanisms. I therefore consider STM and WM approaches only as different views on the same cognitive function and I use the term WM as an umbrella term for active as well as for passive tasks. For example for a prototypical passive WM task is the S1–S2 matching or change detection task in which a stimulus has to be temporarily remembered after it is removed from perception in order to compare it with a shortly afterwards presented test stimulus. The focus of this research is the storage function. Active WM tasks are more heterogeneous but most of them share the aspect that the memory stimulus has to be further processed reorganized, transformed, updated, etc., before it is matched with S2 (see Cornoldi and Vecchi, 2003) for a thorough discussion of the active-passive distinction. Obviously these tasks make it necessary to manipulate the to-be-remembered stimulus. However, all processes operate on specific representations, and therefore even in active WM tasks the storage component is of equal relevance as the processing component. I therefore want to have a closer look on the different types of representations that contribute to visual working memory and to the neural structures that provide them.

The most influential model of WM probably was the tripartite model of Baddeley and Hitch (1974). According to this model, WM consists of three components: the central executive, the phonological loop, and the visuo-spatial sketchpad. The latter two are independent so-called slave systems existing alongside long-term memory (LTM). More recently, the episodic buffer was added as a further component (Baddeley, 2000). The episodic buffer represents integrated multi-modal information from different modalities and also semantic information. The structure of the revised model is illustrated in Fig. 1. As an alternative to separate dedicated memory systems, some authors have suggested that working memory is activated information within LTM (Cowan, 1995, 1999) or activated information in LTM bound to specific context information (Oberauer and Kliegel, 2001). In these models, memory is usually not further subdivided into part systems. A compromise between both positions is the assumption that WM is indeed provided by (active) long-term memory structures, but that some of these structures and processes are domain-specific, meaning that different modalities or different contents (e.g., faces and places) are processed within different brain structures (Fuster, 1997; Ranganath and Blumenfeld, 2005; Ruchkin et al., 2003).

According to the latter view, different working memory performances are provided by networks of control and storage components and it is an empirical question which of these components are domain-specific and which contribute to WM and/or LTM, respectively. In this paper, I adapt the latter position.

The tripartite model encouraged a lot of research on WM (see Baddeley (2003) and Repovš and Baddeley (2006), for reviews). I therefore wish to take this model as a starting point for discussing visuo-spatial working memory (VSWM). I will analyze in what respect VSWM can be considered a unitary part system that is separate from LTM. I will show that it is more reasonable to consider VSWM as a distributed network that processes and represents different types of information than as a self-contained WM system. I will then move on to the topic of visuo-spatial imagery because visual imagination is a prototypical service that is provided by VSWM. Hence, research on visual imagery should give further insights into the function of VSWM and we should expect parallel results in both types of tasks. I will argue that both tasks are provided by highly overlapping (if not the same) neural structures. The main conclusion will be that visual working memory consists of processes governed by prefrontal structures (Courtney, 2004), that these processes operate on different visuo-spatial representations which (b) themselves are provided by the same posterior neural structures that are involved in perception and contribute to long-term memory (Jonides et al., 2005). Some of these structures are modality-specific—they are exclusive to the visual modality—whereas others are domain-specific—they represent specific contents (e.g., spatial information) which are made available by different modalities. From this perspective, visual working or short-term memory is not a service that is provided by a separate system—dedicated only to memory—but it arises as an emergent property of processes running on visual and spatial information represented in a distributed ventral and dorsal network (see Postle (2006) for a similar position). The different compartments are a consequence of the domain-specificity of the involved representations.

1. What is stored within the visuo-spatial sketchpad

The visuo-spatial sketchpad was introduced as an additional short-term store besides the phonological loop. The main argument in favor of another subsystem was the observation that secondary tasks within the same domain (verbal or visual) interfere with each other, whereas two tasks from different domains do not (Baddeley et al., 1999) (see Repovš and Baddeley (2006), for a recent review). However, though it was beyond dispute that verbal and visual working memories should be separated (but see, e.g., Jones et al., 1995), it has always been controversial what should be considered the content of the visuo-spatial sketchpad. Baddeley suggested that the sketchpad is a spatial store. This was mainly concluded from the observation that an auditory-spatial secondary task interfered with a visuo-spatial main task (Baddeley and Lieberman, 1980). The main task was the Brooks matrix task, in which participants are required to remember a path through an imaginary spatial matrix, and the secondary task was auditory spatial tracking. Smyth and Scholey (1994) reported a similar result for the Corsi task—in this task participants memorize a sequence of moves across positions—and an auditory secondary task that demanded spatial attention.1 Because the modalities of the two tasks differ, it was concluded that the spatial and not the visual character of the task must be critical. Contrary to that position, Logie (1986) observed that

\[1\] We will later come back to this task and we will report further results on it.
memory in a visual-imaginal main task was impaired when irrelevant pictures were presented which is a visual and not a spatial interference. In the main task, participants learned a sequence of objects by constructing visual images (the peg word mnemonic). Due to this interference effect, Logie concluded that the visual sketchpad stores both spatial and visual information, but this controversy continued to exist.

For its solution, Logie (1995) postulated that VSWM comprises both types of information that are, however, processed by two different components: a passive visual store (the visual cache) and an active rehearsal process (the inner scribe): the cache stores visual information (e.g., color), the inner scribe stores dynamic information (e.g., movements). Only the latter one is considered spatial, referring to “a representation that involves movement in its broad sense, to incorporate imagined movement as well as physical movement” (Logie, 1995, p. 78). Examples for movements are visual or mental scanning of an image, physical movement to a target, or movement of objects in an array. Support for this distinction is seen in selective interference effects between visual and spatial secondary and main tasks. A prototypical visual task is a short-term visual memory task (e.g., remembering colors or a pattern of filled cells within a matrix (Della Sala et al., 1999); a prototypical spatial task is the Corsi task (see Berch et al., 1998), for review).

Correspondingly, irrelevant visual material (pictures, colors, visual noise) is presented in visual secondary tasks, and spatial tapping is the most frequently used spatial secondary task. The standard result is that combining two tasks of the same type causes lower memory performances than combining two tasks of different types (Della Sala et al., 1999; Hale et al., 1996; Klauser and Zhao, 2003; Logie and Marchetti, 1991; Mohr and Linden, 2005; Tresch et al., 1993; Vuontela et al., 1999). These results clearly demonstrate that in WM visual information is differently processed from spatial information. However, it is still not well specified what these two components are and how they interact. For example, spatially distributed locations (filled cells in a matrix) are considered visual information that is held within the visual cache, and the same locations would be considered spatial when they are part of a sequence of movements that is held within the inner scribe. On the other hand, in a Corsi task, visual qualities – e.g., the symmetry of the spatial layout – also influenced memory performances (Rossi-Arnaud et al., 2006). Given this, the critical feature of the inner scribe might not be its visuo-spatial quality, but its spatio-temporal character (see also Zimmer and Speiser, 2002).

Several years ago, Jones and colleagues already highlighted the relevance of temporal information for the observed effects in verbal working memory (Jones, 1993; Jones et al., 1992). They argued that changing states in the stream of items presented in the interference task (e.g., changing tones) cause cross-talk with the stream of items that has to be remembered in the main task (Jones et al., 1992). Consequently, temporal information and not the modalities of the stimuli were considered as relevant, and the authors therefore searched for similarities between tasks that made it necessary to remember sequences of items in different modalities. In fact they found a number of similarities between serial recall of verbal and spatial material and the interference effect of irrelevant material was strongly reduced if it had no serial component (Jones et al., 1995; Parmentier and Jones, 2000; Tremblay et al., 2006). More recent results suggest that in an auditory spatial working memory task too, temporal information is encoded together with spatial information (Parmentier et al., 2004) and that temporal grouping can structure the path of a Corsi task (Parmentier et al., 2006). Hence, the Corsi task does not only rely on spatial but also on temporal information and results gained with this task cannot unambiguously be attributed to the spatial characteristics of the task.

1.1. What and where

When Logie introduced the distinction between cache and scribe, he drew a parallel between the suggested visual and spatial components of VSWM and the distinction of “what” and “where” pathways in the visual processing stream (Goodale et al., 1994b). The ventral pathway ending in the inferior temporal lobe is considered as relevant for object identification, whereas the dorsal pathway ending in the parietal cortex is considered as relevant for processing of spatial information (Ungerleider and Mishkin, 1982). Numerous dissociations between object and spatial tasks in neuropsychological case studies support that this is a highly relevant division in perceptual processing and in the control of actions (Goodale and Milner, 1992; Goodale and Westwood, 2004). Hecker and Mapperson (1997) were able to show that the distinction has some importance also for stimulus encoding in VSWM tasks. Concurrent chromatic stimulation in the background of a target (processed by the parvo-cellular system and hence in the what-path) impaired color and shape but not location memory, and achromatic flickering (processed by the magno-cellular system and hence in the where-path) behaved in the opposite way. However, though this dissociation demonstrates that the distinction is relevant for encoding, it is less clear whether this result is also a convincing evidence for the existence of two systems in memory. Goodale and Humphrey (1998) discussed that the where-system is used for action, but that it might not be used for memory. For example, patient DF had a lesion in the lateral occipital cortex (LOC) that impaired processing in the ventral stream whereas processing in the dorsal stream was normal—in the experiment it was realized as judging the orientation of a line versus using orientation (of a slit) to guide movements (James et al., 2003). In spite of the impaired “what” pathway she could successfully manipulate objects via the “where” system, but when a delay was introduced between target presentation and execution of the action, she was strongly impaired (Goodale et al., 1994a). From these results it was concluded that the ventral but not the dorsal stream is used for memory. However, recent results with this task with healthy participants suggests that both structures contribute to memory. Ventral structures (LOC) were reactivated while an action was performed from memory though no object was present during performance and these structures did not show enhanced activity during the delay period (Singhal et al., 2006). Structures in the dorsal stream (the anterior intraparietal sulcus) showed also delay activity. I will have a more differentiated look at these structures later.

Differential contributions of ventral and dorsal structures to working memory are also suggested by a number of electrophysiological studies (Ruchkin et al., 1997). These experiments have demonstrated that neural structures underlying object and location memory can be separated in the maintenance phase of working memory tasks. Although this distinction is different from the perception–action dimension investigated in the context of “what” and “where”, it is usually considered as equivalent assuming that object features are processed in the “what” and spatial features in the “where” path. In these experiments, participants are typically instructed to remember a varying number of either objects or locations in an S1–S2 paradigm. During processing, the EEG is recorded and averaged slow potentials during the retention interval are analyzed. Mecklinger and Pfeifer (1996), for instance, presented configurations of simple geometric figures (square, circle, diamond, etc.) and required their participants to look for a figure or a configuration change in object
and spatial tasks, respectively. They observed negative slow waves during maintenance and their amplitudes increased with memory load. In the object task, this effect occurred over the anterior cortex whereas it occurred over posterior parietal and occipital areas in the spatial task. Boschi et al. (2001) presented objects and nonsense 3D bodies in a virtual three-dimensional arrangement, and the target modality was either visual or verbal. With this new material, the topography of the activities was slightly different. In the object task, a positive component was found at occipital sites that they interpreted as a correlate of object processing. In the spatial location task, a negative component was found at parietal sites and additionally a positive slow potential at parietal-occipital electrodes. They concluded that the negative component was related to spatial attention processes whereas the positive component was a correlate of object processing—note that participants’ spatial task was to remember objects at locations.

These data show that neural activities that process visual and spatial information can be separated during maintenance, but they also demonstrate that the topography of the effects vary with the specific demands of the task (content). We will come back to the neural localization of the different VSWM components in the next section. In the present context, it is only of interest that object tasks could be dissociated from spatial ones—leaving open where the object representation is located—and that the spatial tasks were associated with a parietal negative slow potential that the authors related to spatial attention.

1.2. Attention and VSWM

Attention was already mentioned as a critical aspect of spatial WM in earlier work. Smyth and Scholey (1994) suggested that covert shifts of spatial attention are used to rehearse spatial information in the Corsi task. Contrary to that, Klauer and Stegmaier (1997) proposed that not spatial attention but the central executive is involved in the rehearsal process (see also Baddeley, 2003; Fisk and Sharp, 2003; Vandierendonck et al., 2004). This conclusion is based on the observation that non-spatial but attention-demanding tasks also interfere with Corsi tasks. However, other results suggest that although central attention might be a prerequisite for spatial WM tasks, attention has a clear spatial component in this type of task. A statistical argument is that ‘pure’ spatial tasks (e.g., the Corsi task) as well as spatial storage and transformation tasks share variance with central executive measures to a high and similar degree (Miyake et al., 2001). A physiological argument is that the neural structures found active in spatial attention and in VSWM strongly overlap (Corbetta et al., 2002). A direct empirical argument is the observation that spatial rehearsal in WM influences visual perception of patterns at locations that are maintained in memory (see Awh and Jonides, 2001) or Kanwisher and Wojcik (2000) for reviews.

Awh et al. (1998) presented to-be-identified letter-like patterns during the maintenance interval of a spatial WM task. Shorter reaction times were observed when the figures were presented at target locations of the WM task than when they were presented elsewhere in the visual field. In other words, at target locations of the working memory task, visual processing of a stimulus was facilitated very similar to detection tasks in that spatial attention was overtly directed to spatial locations. This facilitation effect due to covert spatial rehearsal was later also demonstrated in an ERP experiment (Awh et al., 2000). Furthermore, when a visual task (e.g., color discrimination) directed spatial attention away from target locations, spatial WM performances were more strongly impaired than when this comparison could be done without spatial reorientation (Awh et al., 1998; Hale et al., 1996). Finally, a sequential spatial task (Corsi) was more strongly impaired by a central executive secondary task (random number generation) than a simultaneous spatial task (Matrix Pattern Task) (Rudkin et al., 2007). Hence, an attention based spatial rehearsal process seems to contribute to (sequential) VSWM, and it might be this mechanism that is interrupted by spatial suppressor tasks like spatial tapping, pointing, tracing moving objects, etc.—see for a similar discussion (Zimmer et al., 2003). This would also explain why a search task during the retention interval—it needs spatial attention—interfered with a spatial change detection task (Woodman and Luck, 2004), but it did not influence performances in a variant of the task with colors and shapes (Woodman et al., 2001).

Another mechanism by that attention contributes to WM is not shifting of attention over spatial locations but directing attention to a memory stimulus which may be an isolated feature or an object. This mechanism has two consequences. One is that directing attention to specific information enhances the neural response to a perceived stimulus in those structures that represent the target. For example, attention to shape or color enhances the activity in extrastriate visual cortices that are specialized for these types of information (Clark et al., 1997; Corbetta et al., 1996). The second component—correlated to enhanced prefrontal activity—is refreshing object information that is held in WM and this process improves memory (Johnson et al., 2004; Raye et al., 2002).

Finally, a third function of attention in VSWM is a gating mechanism (Awh et al., 2006). If we assume that all features in the perceptual field are processed in parallel, but that only a limited number of objects enters visual working memory, a mechanism must be postulated that defines objects and that regulates “entrance” into WM. “Biased competition” (Duncan, 2004) is a possible candidate for this selection. The to-be-solved task sets up specific demands and they facilitate processing of relevant information via top-down influences. In interaction with bottom-up influences (e.g., perceptual salience), this bias favors processing of a limited number of objects, which will therefore “win the race” for representation in WM. According to Cowan’s Embedded Process Model (Cowan, 1988), these objects are in the focus of attention and only about four items can be in this state. This would limit the capacity of WM.

1.3. The capacity of WM

Also Luck and Vogel (1997) and Vogel et al. (2001) estimated the capacity of VSWM as about three to four items (see also Shibuya and Bundesen, 1988). They presented their participants a display showing 3–12 objects for 100 ms, and each “object” was defined by either one unique feature (e.g., a color patch) or by conjunctions of features (e.g., a color, an orientation, and a shape). Nine hundred milliseconds after presentation, a second picture was shown depicting the same array of “objects” either identically or with one item changed. In all conditions, memory was nearly perfect for four items and then it declined. This function was identical for single- and multi-feature objects, and it was even observed for conjunctions of features within the same dimension (e.g., two colors). The authors therefore concluded that VSWM can hold four objects and that each object can represent an arbitrary number of features without any additional costs. Hence, VSWM is not limited in the number of features that can be stored, but in the number of objects.

If the number of objects is relevant, the definition of an object becomes critical. It turned out that this is a difficult task. For example, Vogel et al. (2001) investigated feature conjunctions within the same domain (two colors). They presented a smaller square in the centre of a larger square and both were in different colors. These bicolor patterns behaved as ‘objects’, i.e., the memory
load of four patterns (made up of eight colors) were comparable to four uni-colored patches. However, the authors also mentioned that they had tested a number of alternative two color layouts but only the used arrangement was independent of the number of features. Coloring the two halves of a rectangle in different colors, for example, did not work. In my lab, we also compared memory for conjunctions and features in order to check the reliability of the results (Zimmer, 2007). We measured memory performances for uni-color and bicolor patches of different set sizes using the presentation conditions of Vogel and Machizawa (2004). Additionally, we analyzed slow wave potentials during the retention interval as a function of set size and stimulus condition (feature versus conjunction) because their amplitudes are interpreted as an estimate of processing effort in WM (Vogel and Machizawa, 2004).

With increasing set size memory performances declined and they did so much stronger for conjunctions than for individual features. Obviously conjunctions of features could not be remembered as integrated objects without costs. In the ERP, we observed negative potentials over occipital brain regions that increased with set size. Conjunctions caused somewhat more negative potentials than features, but these costs were not proportional to the number of features. We only got a main effect of stimulus condition. When we measured mental load by the variability of pupil dilatation (Index of Cognitive Activity, Marshall et al., 2004), likewise a main effect was observed. Higher cognitive activity was observed during remembering color conjunctions than colors. This suggests that maintenance processes operate on objects but that processing conjunctions is not for free, it causes additional effort. Furthermore, the observed decrease of memory with increasing set size in the conjunction condition must be caused by a different effect (see the similarity hypothesis discussed below).

Xu (2002) systematically manipulated the way of presentation in a conjunction condition. She observed that the object advantage was highest when the colors belong to the same part of a spatially coherent discrete figure composed of different parts, but an advantage still exists when the colors cover different parts of the same coherent figure compared to covering parts of two spatially separated figures. Delvenne and Bruyer (2004) also presented features in different binding conditions. As we did, and in contrast to Luck and Vogel (1997), they observed that color–color conjunctions caused costs compared to single features even though they belonged to one object. However, conjunctions of features from different dimensions of the same object (shape and texture) behaved like single features probably because they could be integrated as a unitary perceptual object. Later the authors were able to demonstrate that the object advantage was partly due to a reduction of spatial complexity but a residual effect nevertheless remained even if this influence was controlled (Delvenne and Bruyer, 2006). Obviously, it depends on the way the figure is processed in vision whether two features behave as one or two units (cf. also Scholl, 2001).

However, not only the number of objects but also sensory load limits visual working memory. Alvarez and Cavanagh (2004) presented figures of different complexity. They used simple colored squares, letters, line drawings, Chinese letters, random polygons, and shaded cubes. They estimated the capacity of visual working memory as the number of objects that can be stored (see Cowan (2001) for the algorithm). The capacity decreased in the above given order, and the capacity (estimated as reciprocal of 75% proportion correct in a change detection task) was perfectly related \( r^2 = .992 \) to the search rate (ms per item) in a visual search paradigm. Interestingly, for perceptually simple figures, the limit was given by the maximal number of about four objects, but this limit was reduced to about two items for complex figures. However, this is not fully a problem of storage capacity. Partially the reduction was caused by the short presentation time that was too short to encode the to-be-memorized complex stimuli—500 ms in the experiments of Alvarez and Cavanagh (2004). If encoding time was not the limiting factor, the influence of complexity was reduced, but clear residual effects remained. With an encoding time of 3000 ms, the correlation between search rate and VSTM capacity was about .50 (Eng et al., 2005). Therefore, VSTM seems to be limited by two factors: the number of objects and the complexity of each object.

These are clear effects but recent results call their interpretation into question. Awh et al. (2007) argued that complexity does not reduce the number of objects that can be remembered but that it enhances their similarity to distractors and therefore the comparison of the memorized S1 with the presented S2 is more difficult. They again observed that the capacity perfectly correlated with the inverse of the comparison time. Furthermore, even with very complex objects, a cross–category change of features (from Chinese letters to cubes, stimuli that have a low similarity) could be as easily detected as with simple objects whereas a within-category comparison (high similarity) was much harder for complex than for simple objects. Please note that S1 and therefore memory load was always the same in the within- and cross-category condition. According to the authors, this should not happen if complex objects were more often completely lost from memory than simple objects. Furthermore, individual capacities for within-category comparisons highly correlated with each other independent of the item content, and also color and cross–category capacity correlated. However, performances with color did not correlate with within-category capacity and within- and cross-category capacity did not correlate. The authors concluded that the capacity has two components: one is a storage capacity that limits the number of to be remembered objects independent of complexity, and one is the efficiency of the discrimination process for complex material. We will see later that these factors may come to effect in different neural structures. Some parts of parietal cortex were only sensitive for the number of objects (inferior intraparietal sulcus) and others (superior parietal and lateral occipital regions) were sensitive for visual complexity (Xu and Chun, 2006). This suggests that it may be preferable not to speak of storage and processing components but of representations of an object at different levels of abstraction and the efficiency effect is a consequence of the ease of comparison at the level of abstraction being suitable for the detection of a stimulus change.

However, VSTM is not solely limited by the sensory complexity of a stimulus defined by its perceptual characteristics but more so by the perceived complexity which depends on the interaction between the observers pre-knowledge and the complexity of the stimulus. I therefore want to make a distinction between exogenous and endogenous complexity. Exogenous complexity is defined by geometric features of the stimulus. Examples are number of stimulus parts and their spatial distribution, spatial frequencies of the object–specific details, symmetry or regularity of the figure, number of strokes, etc. Endogenous complexity is the perceived complexity for the observer which depends on the availability of long-term memory entries for the stimulus and the observer's expertise with this type of material. A very complex character can be rather simple if the perceiver has a long-term memory entry that immediately represents the input and chunks the perceptual features. This is possible because VSTM does not store raw input but it represents “interpreted” visual input making use of pre-existing perceptual knowledge, e.g., visual prototypes, characters, etc. We tested this hypothesis by comparing VSTM capacity for color patches, Chinese characters, Chinese pseudo characters and Chinese nonsense characters in an S1–S2 paradigm (Zimmer and Fu, 2008). S1 was a bit longer (1000 ms) than in the
reported studies to reduce the effect of encoding time on complexity. Chinese characters are exogenously more complex than color patches, and we additionally varied exogenous complexity of the characters by presenting characters with different numbers of strokes. However, for literal Chinese native speakers the perceived complexity of characters should be strongly reduced because they have long-term entries for these stimuli. Therefore, the three types of Chinese characters should have different endogenous complexity for Chinese and German native speakers. For Chinese, the endogenous complexity increases from real to nonsense characters whereas for Germans, only the distinction between colors and characters makes a difference. Working memory capacity (k) for color patches was the same for German and Chinese participants. For Germans, capacity for all Chinese characters was very low independent of the type of character but dependent on endogenous complexity. For Chinese participants, capacity for characters was nearly as high as that for color patches, and it decreased to pseudo characters and further to nonsense characters. Furthermore, when no long-term memory entry existed (nonsense character) a clear effect of the character’s exogenous complexity was observed. The same happened when real characters were sufficiently complex (about 15 strokes). This clearly demonstrates that both factors – the item’s exogenous complexity and the observer’s expertise – influence working memory performances and that finally WM capacity is a function of endogenous complexity.

1.4. Individual differences in VSWM

So far, we have seen that stimulus qualities limit performances in VSWM. Another factor that does so is the efficiency of a participant’s memory. Eng et al. (2005) reported strong inter-individual differences in VSWM capacity. The capacity (calculated as the number of items that can be held in memory) ranged from 1 to 3. However, though participants’ capacities strongly differed, these differences were not influenced by item complexity, and the capacities highly correlated across different stimulus types (r = .82). This suggests a limit in the number of objects. The authors also reported that capacity was not influenced by practice. Olson and Jiang (2004) observed a similar result. Stimuli in which participants had been trained were not better remembered than new ones. In another study, participants’ performances indeed improved but this was a general practice effect because learning was not restricted to the trained items (Chen et al., 2006). Olesen et al. (2004) trained participants in a Corsi like task and they observed increased performances and enhanced neural activity in middle frontal, superior and inferior parietal cortices. Moore et al. (2006) showed similar effects for an object WM task (training in novel object categories) in prefrontal and occipital-temporal cortices. Furthermore, the above presented effect of endogenous complexity can also be considered as an effect of a very long training. The reason for the different results are not yet known. Obviously, some individual characteristics set a limit to the number of objects that can be stored in VSWM but the mechanisms of this limitation and its plasticity is not yet understood. I will discuss neural processes that may be suitable to explain this limitation later.

In recent research, another phenomenon was reported that can be either a reason for the limited individual VSWM capacity or a correlate of the capacity limitation of neural processes (Vogel and Machizawa, 2004). The authors were interested in the electrophysiological correlates of interindividual differences in VSWM. They presented a number of color patches in the left and right hemifield. Briefly before stimulus presentation, they indicated which hemifield constitutes the memory array. One second later, the target stimulus was shown for comparison. During the delay period, negative ERPs were observed over the contralateral occipital parietal cortex (contralateral delay activity) (see also McCollough et al., 2007). This effect was already observed before in a study by Klaver et al. (1999). The new observation was that the amplitude increased with the number of items and this increase was highly correlated with participants’ memory capacity. When the number of items approached individual memory capacity, the negative potential reached its asymptote. More than this, the memory capacity is highly related to the filtering efficiency. This is the efficiency to ignore irrelevant information and it was measured as the difference of contralateral amplitudes when two isolated items were remembered and when two items in the context of two irrelevant items were remembered divided by the maximum amplitude difference (the difference between two and four items) (Vogel et al., 2005). Participants with high capacity seemed perfectly to filter irrelevant items whereas irrelevant items seemed to enter VSWM of low-capacity participants where they competed for the limited ‘storage space’ (see Cowan and Morey (2006), for a discussion).

The ability to inhibit irrelevant information may be a general ability and it may also explain the correlation between WM capacity and general intelligence (Engle et al., 1999). Participants with higher capacities seem to be better in directing their attention to relevant information than those with lower capacities. An empirical argument in support of the assumption that working memory capacity has something to do with the efficiency of selection and controlled attention (Engle, 2002) is the observation that participants with high capacity performed a voluntary attention shift faster than those with low capacity although they did not differ in automatic attention shifts (Kane et al., 2001). Furthermore, in a demanding spatial WM task, the power of the EEG in the theta range (4–7 Hz) showed a strong increase with practice in high capacity subjects that was not found in low capacity subjects (Gevins and Smith, 2000). This theta component observed at frontal electrodes is considered as a signature of focused attention (Gevins et al., 1997).

1.5. Features and conjunctions

Unfortunately, the maximum number of objects that can be remembered is only a part of the unsolved issue. It is even all but clear that objects are the units. As already reported, it was not always observed that WM is a function of the number of objects and not of the number of features. Besides others, Wheeler and Treisman (2002) could also not replicate the result of Luck and Vogel (1997). In their studies, memory for three bicolored objects (i.e., six color features at three objects) was identical to memory for six objects with one feature each, and for both item types, memory was worse than memory for three objects with one feature each. This suggests that the units are not objects or, if they are objects, that nevertheless binding features within objects (i.e., knowing that features belong to the same object) causes costs—at least if the features belong to the same dimension.

In order to test the binding costs, in further experiments the authors compared feature binding across different dimensions with memory for single features. In the critical binding trials, a distractor item was a recombination of old features of different objects so that remembering individual features was not sufficient to solve the memory task. Therefore, participants had to remember the specific combination of features in order to make a correct rejection. Treisman and Wheeler observed lower memory performances in the binding than in the feature condition, but this effect was modulated by the way memory was tested. If at test the whole display with all objects was shown, memory performances in the
binding condition was lower than memory in the feature condition. Interestingly, however, when only one object was presented as test stimulus, memory for the correct binding of features was as good as memory for the single feature that had the lowest memory performance. Pure feature memory was not influenced by this manipulation.

From these and other results (see Treisman (2006), for a recent summary), Treisman and her colleagues have drawn several conclusions about the properties of VSWM. During perception, attended objects are represented as object files (Treisman and Gelade, 1980). VSWM stores a limited number (two or three) of object files. Under a similar perspective of scene perception, also Hollingworth (2004) assumed that VSWM holds about two object files. Object files bind the stimulus features that are represented in independent, multi-dimensional feature stores, each with its own limited capacity (see Zimmer et al. (2007) for a discussion of this type of sensory memory). According to Treisman, feature binding needs attention and the binding process is vulnerable. Object files are updated if the input changes, and this overwrites former conclusions about the properties of VSWM. During perception, object files bind the stimulus features that are represented in independent, multi-dimensional feature stores, each with its own limited capacity (see Zimmer et al. (2007) for a discussion of this type of sensory memory). According to Treisman, feature binding – as Treisman assumed – or because of some other unsolved issue whether this is because attention is necessary for binding – as Treisman assumed – or because of some other mechanism. For example, a demanding secondary task did not influence memory for bound features more than for individual features, as long as the items are simultaneously presented (Allen et al., 2006). This suggests that binding has no specific demands on attention. A similar conclusion follows from the study of Gajewski and Brockmole (2006). They applied an exogenous cue during maintenance of bound features that should distract attention. The cue was efficient because it enhanced memory at target locations, but it did not selectively impair conjunctions. In invalid trials, participants recalled either both features of the conjunction or none. These results speak in favor of the assumption that VSWM holds bound information and that attention is not necessary to maintain binding. However, it is possible to focus only on isolated features and not to memorize bound objects as Woodman and Vogel (2008) could recently show.

Considering the variability of these findings, we have to admit that feature binding is neither sufficiently specified in WM models nor is the mechanism of binding really understood (e.g., Walker and Cuthbert, 1998; Patterson et al., 2007). According to my view, it is necessary to integrate ideas from models of object perception into WM to accomplish an adequate understanding of binding and WM. For example, the distinction between integral and separable dimensions as discussed by Garner (1974) is probably relevant here and also the one between different independent processing streams in object identification (Cant and Goodale, 2007). A consequence hereof would be that object representations are considered more fragmented and distributed than it is suggested by speaking of object files. Perceived objects can have multiple representations that partially depend on the perceiver’s intention and partially on perceptual qualities of the stimulus and these fragments can be bound in various ways. Another consequence is that processing of locations has to be reconsidered.

Locations are separable from visual appearance and they can have different roles dependent on the task. On the one hand, a location can be a characteristic feature that individuates a specific object among a number of identical objects. Corst tasks usually are of this type because perceptually identical objects mark locations. In these tasks, location is an individual feature that has not to be bound to solve the task. The memory “address” is the location. On the other hand, when a distinct object is perceived, its location is an arbitrary feature that is not integral to the object like color may be. In this case, object features are dominant and locations are part of the binding process. Now an object file may be the memory “address”. Treisman tackled the latter situation. According to her model, when attention is directed to an object presented in a spatial location, an object token is generated that binds features. Location is the joint information shared by all features and it is represented in a master map of locations. This map might be provided by the parietal lobe (Friedman-Hill et al., 1995; Shafritz et al., 2002). Hence, location and object information are separable features but nevertheless there is a high probability that both are spontaneously bound if objects are attended, and both are useful cues for each other (Treisman and Zhang, 2006). Two notes are indicated here. First, many maps exist in the cognitive system, e.g., different types of retinotopic representations and different types of coordinates. The master map is related to spatial information in terms of a conscious representation of environmental space. Second, locations of objects can serve another function. They constitute configurations and this will also make spatial information available in WM as I will discuss later.

1.6. Spatial information reconsidered

In light of these data, we should revise our opinion on the components of VSWM. We should distinguish three separate representational formats: a visual representation of perceptual information of objects (its appearance), spatial coordinates as to where objects are located, and object files representing bound information. Visual representations therefore exist at different levels, as individual features (appearance) and as bound object files. Independently, a representation of spatial locations (coordinates) exists, which can be used by spatial attention and which represents the individual’s spatial environment. Finally, because both types of information are separated, an additional process binds objects to locations. We can also distinguish three different memory tasks: object memory, spatial memory, and memory for object to location assignments (Postma and De Haan, 1996). From this perspective, spatial WM for objects is neither visual nor spatial, but a compound that relies on object files.

Consequently, according to the revised tripartite WM model (Baddeley, 2000), information on the location of objects should neither be held in the visual cache nor in the inner scribe, but in the episodic buffer. As mentioned, this buffer should represent integrated multi-modal information from different systems and modalities. It can in principle be considered a specific state of (episodic) LTM, in which recently processed information is maintained after it has been integrated as it is assumed in unitary memory models. Support for this assumption is the observation
that neither a visual nor a spatial interference task during maintenance influenced spatial memory, whereas factors known to be relevant for LTM did. For instance, Zimmer et al. (2003) demonstrated that additional visual input during maintenance and also spatial tapping did not impair location memory, and spatial WM (re-positioning) performance was superior for real objects as compared to visually distinct but arbitrary shapes (i.e., meaningless material). The influence of the material’s meaningfulness on memory performance suggests a contribution of LTM to this spatial WM task.

Additionally, we have collected data supporting that spatial working memory is provided by a domain-general structure as the episodic buffer should be. Spatial WM seems to represent information independently of stimulus modality which is why we have suggested a common coding in spatial WM (Lehnert and Zimmer, 2006). We presented visual and auditory “objects” in a spatial WM task (pictures versus natural sounds of objects), and we did so in modality-pure or in mixed lists—half of the items were pictures and half of them sounds. One would expect better memory performances in mixed lists than in pure lists if separate modality-specific stores exist, but no effect of list structure if a common coding exists. Spatial short-term memory decreased with increasing list length, the costs per item were the same for visual and auditory objects, and memory in modality-pure lists was as high as memory in mixed lists, in which half of the items were presented as pictures and half as sounds. This is speaking for common processes coding spatial information in both modalities.

A comparable result was observed by Saults and Cowan (2007) who attributed the equal performances to central capacity (for the moment we ignore the modality-specific effects). Similarly, in an fMRI study, Arnott et al. (2005) observed activations in the parietal cortex if locations of sounds had to be memorized similar to those of pictures. In a comparable result was observed by Saults and Cowan (2007) who attributed the equal performances to central capacity (for the moment we ignore the modality-specific effects). Similarly, in an fMRI study, Arnott et al. (2005) observed activations in the parietal cortex if locations of sounds had to be memorized similar to those of pictures. In a comparable result was observed by Saults and Cowan (2007) who attributed the equal performances to central capacity (for the moment we ignore the modality-specific effects). Similarly, in an fMRI study, Arnott et al. (2005) observed activations in the parietal cortex if locations of sounds had to be memorized similar to those of pictures. Moreover, a shape-like recognition of sounds made at the time of presentation did influence spatial memory, whereas factors known to be relevant for LTM did. For instance, Zimmer et al. (2003) demonstrated that additional visual input during maintenance and also spatial tapping did not impair location memory, and spatial WM (re-positioning) performance was superior for real objects as compared to visually distinct but arbitrary shapes (i.e., meaningless material). The influence of the material’s meaningfulness on memory performance suggests a contribution of LTM to this spatial WM task.

However, in a recent series of experiments we could show that explaining the effect by a mismatch in the visual cache is probably not correct (Zimmer and Lehnert, 2006). The spatial mismatch effect was not affected if additional visual input was processed during the maintenance interval, and it even occurred if the items’ format was changed from study to test—objects were shown at study and their names at test, or vice versa. In contrast to this insensitivity to sensory manipulations, the mismatch effect was sensitive to figurative qualities of the layout. The size of the spatial mismatch effect varied with qualities of the global gestalt of the layout. Trials with an identical S2 were fastest. Changing absolute locations but keeping the viewpoint-specific global shape of the layout constant caused a small increase in response times (e.g., presenting objects in an L-shaped configuration as S1 and expanding the whole L-shaped configuration from study to test). Rotating the global figure as a whole caused a higher cost in response times (still presenting the objects in an L-shaped configuration but once upright and once tilted). Finally, presenting objects in a completely new configuration led to the longest response times. The authors took this as evidence that the spatial mismatch effect is in fact a “shape” mismatch effect. The locations of the objects define a global “shape” (the corners of a polygon) that is remembered as a global figure similar to a geometric object. Verbal labels given to the configuration were rejected as possible reasons for these effects because the same data were observed for configurations with a bad gestalt that could not be named. Other observations also suggest that locations of objects can be represented as a configuration and that this “global shape” influences WM. In a change detection task, for example, detection of location changes is not influenced by changed object features (colors or shapes) but changing features that influence perceptual grouping of objects impaired the detection of changed locations (Jiang et al., 2004). Similarly, absolute as well as relative locations influenced performances in a cueing paradigm (Jiang and Wagner, 2004).

From these results it follows that locations of objects are the origin of two types of spatial representations. They define a visual configuration (probably in object-centered coordinates) similar to a shape, and they define spatial coordinates in the environment. The latter is the (supra-modal) representation that has been discussed before and these locations serve as target locations for the direction of spatial attention and movement (probably in egocentric and body-centered coordinates). This representation is also used for spatially directed actions, including overt motor movements (e.g., grasping or pointing). The distinction between these two types of spatial information is related to Logie’s distinction between spatial information in the visual cache and the inner scribe (Logie, 1995). However, it is not the same. In contrast to his position, the static–dynamic distinction is not critical. The important aspect is that two forms of spatial representation exist which probably have different frames of reference. The visual system represents the shape and visual appearance of objects and it can represent “spatial” information as configuration in a shape-like manner—one may think of grouping...
of low spatial frequency information. The multi-modal spatial system represents unique spatial locations that map to the environment and can therefore be used to direct attention and to guide actions.

According to my view, therefore, the spatial system is not specifically dedicated to dynamic information. The specific interference that dynamic information has on spatial tasks is probably only caused by the fact that moving objects capture spatial attention and that spatial attention is continuously (re)directed in order to “follow” the object. This impairs spatial rehearsal. The dynamic–static characteristic of a stimulus per se does not qualify a specific task as spatial or visual. Our recent experiments on working memory for biological movements have brought direct evidence for this position. We presented point light walkers (Johansson, 1973) in an S1–S2 WM task, and we assigned different types of interference during the retention interval (Zimmer et al., submitted for publication). Memory was not specifically impaired by spatial suppressor tasks. Spatial tapping had only a small negative effect on memory, and this effect was of the same size as non-spatial tapping. Because the interference effect was independent of the spatial quality of tapping, we concluded that the interference is only a consequence of performing an action and not of processing spatial information. On the other hand, all kinds of visual input clearly impaired WM for dynamic stimuli, and even static visual input did so. Presenting unattended color patterns during maintenance had a negative effect, and presenting irrelevant walkers interfered somewhat more. Obviously, there is an effect of distracting attention and additionally an independent effect of sensory similarity between the stimuli of the main and interference tasks. The most parsimonious way of explanation is the assumption that also point light walkers are represented as object files, integrating dynamic information of movements and these features are handled like other types of visual features by the ventral system even though – as we will later see – different neural structures within this path are involved.

In anticipation of this discussion, however, I already want to mention a result from our lab that provided direct evidence for the assumption that spatial (location) and dynamic information is differently processed in WM. We presented dynamic spatial information (a moving dot) in an S1–S2 matching task, and we analyzed neural activity during an extended maintenance interval (Umla-Runge et al., submitted for publication). At the beginning of the empty retention interval a cue indicated whether in this trial movement (velocity/trajectory) or location information (the end position of the dot) is relevant for S2 comparison. Regions that are also involved in movement perception (especially MT/V5) were differentially activated in the movement condition, whereas regions that are related to spatial processing (e.g., parahippocampal and parietal regions) showed enhanced activity in the location condition. This clearly shows that even with dynamic stimuli location and dynamic information are processed differently in WM.

1.7. Interim result

When we summarize the results on VSTM that we had discussed so far, a structure is visible that has some commonalities with the tripartite WM model and others with a unitary model. It is an advanced tripartite model that is best described as a distributed domain-specific sensory and modality-unspecific spatial working memory. A sketch of the model is illustrated in Fig. 2.

We still have domain-specific (e.g., different types of visual and spatial) information. However, it is not assumed that they are represented in separate buffers different from structures used in perception and LTM. In general, memory should be provided by distributed representations, comprising different types of domain-specific modules that process and represent specific types of sensory input. WM as well as LTM use partially the same structures. Active representations appear like independent buffers because they are accessed by different mechanisms than deactivated memory entries.

Within WM one can distinguish different types of visual information (appearance) in the ventral stream – e.g., shape, color, biological movement, visual-dynamics – and spatial information in the dorsal stream. Further non-visual domains represent phonological and auditory (sound) information and this is not an exhaustive list. Each specific type is processed by a content-specific module. Item representations are hierarchically organized from individual features to object files that bind features belonging to the same object. Different types of binding are likely for features within an object and those between objects or between different modalities.

Some of these modules are modality-specific because they code information from a single modality, but other modules are thought to be supra-modal, as, for example, the representation of environmental space. On these representations, specific processes operate. Hitherto, I have only discussed attention operating on spatial representations; some ideas about other processes will be discussed later. Focusing attention to specific content (a location

![Fig. 2. An illustration of the domain-specific sensory and modality-unspecific spatial components of working memory.](image)
or an item) is probably the mechanism that realizes active maintenance in WM.

2. Neural structures involved in VSWM

I have argued in favor of the assumption that VSWM is not a monolithic component but it consists of different part processes operating on different representations. The visual component of VSWM represents the object’s visual appearance at different levels – from isolated features to integrated objects – within domain-specific structures and spatial configurations. Independent of that, at a supra-modal level, the spatial component represents spatial positions that are also used for directing spatial attention and movements. Although reference was made to neurocognitive results already on various occasions in this paper, in the following I want to have a more systematic look at the neural structures that might provide the different performances.

Since the advent of brain imaging research, many studies were conducted to reveal these structures. Among the first experiments to systematically investigate visual and spatial WM were the PET studies of Smith and Jonides (Jonides et al., 1993; Smith et al., 1996). In these studies, letters, dots, or nonsense figures were presented as stimuli for verbal, spatial, or visual WM S1–S2 tasks, respectively. The retention interval was 3 s and the reference stimulus was either one item back or two items back. The two-back condition was considered an active task, because WM had to be continuously updated. The authors came to the conclusion that verbal WM is provided by the left and spatial WM by the right hemisphere, and that the storage component is localized more posterior, whereas maintenance processes are localized in frontal brain regions. Object memory partially activated the same structures (parietal and left frontal), and additionally the left inferotemporal cortex (see Smith et al. (1996), for a review).

In Henson’s review (Henson, 2001), similar areas were mentioned and he explicitly related them to the components of the tripartite VSWM model. He located the visual cache in the left inferior-temporal cortex and bilateral in the anterior part of the occipital lobe; visuo-spatial components were located within the right hemisphere. The inner scribe was assigned to the right superior parietal area (BA 7). This area was seen as part of a network implementing the spatial rehearsal process together with right premotor (BA 6) and right inferior frontal areas (BA 47). However, in a more recent review this dominance of the right hemisphere was not substantiated (Wager and Smith, 2003). In most experiments on VSWM the activation was bilateral. Only in verbal tasks a lateralization to the left was observed, and it increased with task difficulty. Additionally, the differentiation between “what” and “where” was generally supported. Spatial tasks activated the dorsal path, whereas object tasks activated the ventral path.

2.1. The prefrontal cortex in WM

Other experiments have suggested a more differentiated picture. Due to the results of these studies, the prefrontal cortex (PFC) got a more prominent role, because during the delay period this part of the brain was active even when distractors were presented during maintenance, whereas some posterior neural structures lost their activity (Constantinidis and Steinmetz, 1996; Miller and Desimone, 1994). A contribution of the frontal lobe to WM is beyond dispute; however, it is controversial how the PFC realizes this function, and especially whether it is also a storage site of WM. Some authors have suggested that the distinction between “what” and “where” known from the posterior cortex has a correspondence in the frontal cortex. According to this view, the dorsolateral PFC maintains spatial information, whereas the ventrolateral PFC provides object information (e.g., Sala et al., 2003; Ungerleider et al., 1998; Ventre-Dominey et al., 2005; Wilson et al., 1993). Ventre-Dominey et al. (2005), for example, had their subjects solve complex WM tasks. For the spatial task, they found a dorsal pathway between dorsal prefrontal and parietal-occipital cortex predominantly in the right hemisphere. The non-spatial WM task activated a ventral pathway between temporo-occipital cortex and the ventral PFC predominantly in the left hemisphere.

Others suggested that the frontal components do not differ in their contribution to visual and spatial WM, but in the type of processes they perform (Owen, 1997; Petrides, 2005). Whereas the more ventrolateral PFC (VLPFC; BA 45/47) should be mainly involved in maintenance and retrieval of recently encoded information, the dorsolateral part of PFC (DLPFC; BA 9/46) should contribute to monitoring and active processing (Curtis and D’Esposito, 2003). Several results can be quoted as evidence for this position. In different modalities and independent of the object or spatial characteristic of the task, enhanced activity of DLPFC was found in tasks that require high monitoring and active restructuring. Examples are reordering, resolution of interference, target selection, updating of information, and coping with high memory load (Bor et al., 2001, 2003; Constantinidis and Wang, 2004; D’Esposito et al., 1998, 1999; Postle and D’Esposito, 1999; Postle et al., 2000a,b; Rypma, 2006). Furthermore, activity in these brain areas varied with the regularity of the spatial grid and task difficulty, probably because this variable influences the relevance of strategic components for working memory (Bor and Owen, 2006). Interestingly, the DLPFC was also more active if serial order was relevant (Amiez and Petrides, 2007), a specific demand that we considered as critical in the Corsi task—however, note that other studies in order tasks found specific activity in the posterior parietal cortex (Marshuetz et al., 2006). In general, the DLPFC seems to be more important for higher cognitive control and monitoring tasks, whereas in contrast, activity in the ventro-lateral cortex varies with retrieval demands as it was found in many studies of Petrides and colleagues (Champod and Petrides, 2007; Owen et al., 1996; Petrides, 2000, 2002, 2005). I therefore agree on the latter view. The frontal cortex has an executive function – and in this function it is relevant for WM – whereas the posterior cortex is the storage site of the specific information, see also Fuster (1997). The different subregions of the PFC along the ventral-dorsal axis very likely follow a functional differentiation: maintenance/retrieval versus active control.

2.2. Posterior structures in WM

In line with this interpretation, at the neuronal level we have so far distinguished structures in the frontal lobe that make available passive and others that make available active processes during maintenance, and we have suggested that posterior areas store the encoded information. What remains to be presented are more specific results demonstrating which posterior structures provide spatial and object information, respectively. Experiments contrasting these two tasks have yielded consistent results. Spatial tasks showed activity in the parietal cortex, whereas object tasks showed activity in the inferotemporal, lateral temporal-occipital, and fusiform cortex (Druzgal and D’Esposito, 2003; Hautzel et al., 2002; Postle et al., 2000, 2003; Ranganath et al., 2004a,b; Rolke et al., 2000; Ventre-Dominey et al., 2005). At the same time, lesions of these structures impaired specific WM tasks. For example, Schoppig et al. (1999) demonstrated impaired color
memory after lesions in inferior occipital, fusiform, and parietal-occipital areas.

The dissociation between object and location information is also not restricted to visual input. Auditory input showed a similar differentiation. Rämä et al. (2004) observed parietal activation in a visual as well as in an auditory location task. Arnott et al. (2005) reported a parietal activation in a localization task with sounds, whereas identification of sounds activated the rostral superior temporal lobe. Similarly, Rämä et al. (2004) showed that the intraparietal sulcus was active in a sound localization task, whereas the right superior temporal sulcus was active in an identity task. Altmann et al. (2007) observed for pattern changes of sounds activity in the anterior–superior temporal gyrus, whereas the posterior–superior temporal gyrus and the planum temporale were more active in location tasks. The latter structure was also found active in another study if sound localization was not the main task but it was ignored because participants attended to the visual modality (Deoull et al., 2007). It is likely that planum temporale is a primary area for localization of sounds and parietal structures represent (supra-modal) locations in the spatial environment. Even tactile versions of WM tasks yielded a similar picture: the primary sensory cortex contributed to feature memory (Harris et al., 2002), and the parietal cortex was active in a spatial version of the task (Ricciardi et al., 2006).

Obviously, modality-specific mostly posterior areas are those that process specific types of information. Furthermore, these structures are the same that were found to be active in the respective long-term memory tasks (Garoff et al., 2005; Haxby et al., 1994; Khader et al., 2005; Rösler et al., 1995). Ranganath et al. (2004b) provided direct evidence for this assumption in a within subject study. They identified regions in the fusiform face area and in parahippocampal place area that showed category-specific activity and demonstrated that these areas were also active in a WM task with the respective stimuli. PFC activity, in contrast, was modulated by memory load and not by modality. Transcranial magnetic stimulation (TMS) could additionally support that the activities in the posterior structures have a causal function for the memory tasks (Koch et al., 2005). When the temporal cortex was stimulated during the maintenance interval, responses were selectively slowed in object tasks, whereas stimulating the parietal cortex caused slower responses in spatial tasks. TMS at DLPFC worsened performance in both tasks, and TMS at the superior frontal gyrus impaired spatial memory (see next paragraph for an explanation of the latter finding). Taken together, these data suggest that the posterior cortex is the storage site of WM. Location information is represented in the parietal cortex, independently of input modality, whereas sensory information is represented and maintained in the sensory association cortices of the respective modalities and features.

Interestingly, in delayed-saccade tasks, a form of spatial memory could be isolated that was not tied to the posterior but to the frontal cortex (for a detailed review see Curtis, 2006). In these tasks a target location is marked and after a maintenance interval the eyes have to be moved to this location. In addition to the intraparietal sulcus (IPS), areas of the dorsal prefrontal cortex (BA 8) were usually active in this task (e.g., Brown et al., 2004; Courtney et al., 1998; Sweeney et al., 1996; Zarahn et al., 2000). These areas are part of or connected to the frontal eye field (FEF) and it was therefore argued that the activity is a correlate of “covert motor articulations” (Curtis and D’Esposito, 2003). The dorsal PFC might represent oculomotor coordinates for eye movements, and in this way it may contribute to spatial memory. Therefore, we might have two representations providing spatial WM. A retrospective code represents the past event as a memory trace of previous input in posterior (parietal) brain regions, and a prospective code in the dorsal PFC represents coordinates for future actions (eye movements). Curtis et al. (2004) were able to substantiate this by contrasting two spatial WM tasks. In the matching condition, eye movements could be planned before the maintenance interval, whereas in the non-matching condition, they could only be planned at the end of the interval so that participants had to rely more on sensory information. Correspondingly, the matching condition showed higher activity than the non-matching condition in the IPS, whereas this relation was reversed in the intraparietal sulcus. In both areas, cells were found with activity corresponding to participants’ performances. Coherence analysis confirmed the existence of these two networks for spatial WM (Curtis et al., 2005).

However, a more detailed analysis reveals that the posterior structures (IPS and the parietal cortex) can even be further subdivided into smaller regions specialized for different functions (Astafiev et al., 2003; Greffkes and Fink, 2005). The anterior part of the intraparietal area seems to be involved in processing 3D structure and orientation of objects during fixation and manipulation. It is specifically active in cross-modal object recognition (e.g., recognizing an object by tactile manipulation). The medial IPS is involved in visuo-motor coordination of hand movements and targets including pointing. The ventral area of the IPS is part of a network including the posterior parietal cortex that processes motion and spatial information from different modalities and represents location in different frames of reference—including a body-centered representation (Andersen et al., 1997). Hence, this part of the parietal cortex is a good candidate for a supra-modal representation of space that allows cross-modal spatial coordination for motor actions and direction of attention. This functional behavior of the parietal cortex is in good agreement with the reported results from studies on VSVWM.

2.3. Sensory features in the posterior cortex

After discussing brain regions that process spatial information, I want to have a look at representations for sensory features. Yoon et al. (2006) ran an fMRI study using an interference technique to investigate the corresponding network for maintenance of visual material. Subjects’ main task was to remember faces, and either faces or scenes were presented during maintenance. The lateral PFC and the visual association cortex were active during maintenance, and activity was reduced in both structures by the presence of respective additional material. Only the PFC, not the association cortex was selectively impaired, and functional connectivity between both areas was disrupted. The authors concluded that the PFC codes abstract mnemonic information that is represented in the posterior areas—one might think of a pointer structure. However, it would be wrong to interpret this result in favor of a position stating that the PFC is the true residence of WM. Single cell recordings with animals have shown that posterior areas can also be selectively impaired. For instance, in a study by Bisley and Goldberg (2003), cells in the lateral IPS were active during the delay of a delayed-saccade task. Higher neural activity during maintenance was associated with an advantage in identification of objects presented at target locations (see the studies of Awh and colleagues reported above); presentation of a spatial distractor increased firing of neurons representing the distractor location, and it also reduced the perceptual advantage at target locations.

Song and Jiang (2006) directly investigated the neural structures that provide WM for visual information. They were interested in the influence of the number of objects (load) and of the visual complexity on neural activity. For that purpose they contrasted memory for color (simple feature) with memory for...
random polygons (complex shape), and memory for conjunctions of both features (complex conjunction) in different set sizes. In the superior parietal cortex activity increased with set size, and the two complex conditions showed higher activities than the simple color condition (the two complex conditions were not different). The lateral occipital cortex showed the same dependence on feature complexity but it was not dependent on load; the inferior frontal sulcus showed the reversed pattern—load but no feature effects. Ikeda and Osaka (2007) compared WM for color words and colors that could be discriminated only by sensory information (slight hue differences) or also by color names. The ‘verbal’ conditions showed stronger activations in left inferior frontal, left premotor and left parietal areas, whereas the visual conditions showed stronger activation in the right inferior frontal gyrus and intraparietal sulcus. Similarly, Todd and Marois (2004) observed that activity in IPS and intracerebral sulcus increased with the number of items in WM, and the signal change correlated with the capacity of WM (estimated by the k index). The same authors could later show that the increase in activity from set size 1–3 correlated with the individual maximal storage capacity ($K_{max}$) (Todd and Marois, 2005). Recently, Xu and Chun (2006) contrasted neural WM activity for features of different complexity and for different numbers of objects. They reported that the inferior IPS seemed to be able to store about four objects independent of their complexity whereas activity in the superior IPS and the lateral occipital cortex depended on both aspects. The latter structures represented fewer than four objects (activity reached its asymptote) as item complexity increased.

Data suggest that VSWM is not a monolithic component that stores units of holistic objects. On the contrary, features of objects are stored in a distributed network that represents different aspects of objects, so that multiple visuo-spatial representations seem to exist in the posterior cortex. Some of them treat objects as units independent of the features, others code individual features and therefore they depend on the features’ complexity. Additionally, structures in the PFC contribute to WM and it is likely that they have a retrieval function relevant for the maintenance of information in WM.

2.4. Neural correlates of feature binding

However, even this more complex model is not yet sufficiently differentiated. In the presented experiments, usually spatial information was the location of an isolated (nonsense) figure or of dots, and object information was memory for this figure or for a visual feature like color. When presenting the behavioral data, we discussed that objects and locations (and different features) have to be bound. We have not yet discussed which neural structure(s) might achieve this binding of information (see the contributions in Zimmer et al. (2006b) for a comprehensive discussion of neurocognitive aspects of binding).

Several structures were mentioned as relevant for binding, ranging from PFC to posterior and mediotemporal areas. Mitchell et al. (2004) presented pictures or words on the left or right side of a computer monitor and instructed their participants to decide whether the test item is old or new, which modality the study item had, or whether it was presented on the same side of the monitor during study. In the latter two conditions, memory for item-context binding was necessary. The dorsolateral PFC (BA 9) was more active in these binding conditions as compared to pure (old/new) item recognition. However, also other brain areas were considered as functional for binding. Munk et al. (2002) discussed that the secondary motor area and the anterior cingulate cortex are specifically involved in binding processes. They also emphasized that conjunction tasks are not simply the sum of the constituents’ activations, although brain areas that process the individual features were always active in the conjunction conditions. Others considered different prefrontal structures (BA 6, 32) as relevant, based on studies contrasting feature with conjunction memory in elderly and younger people (Mitchell et al., 2000) (see also Prabhakaran et al., 2000). However, the reversed result – posterior, but no PFC activity – has also been found. Shafrizt et al. (2002) considered the parietal cortex as critical for feature binding. When the conjunction of color and shape was relevant the IPS was more active than when only color was critical. The authors explained this by the importance of spatial attention for binding (see also the discussion of Treisman’s position). As already mentioned, Song and Jiang (2006) observed that in superior parietal and lateral occipital cortex memory for the conjunction of color and shape caused more neural activity than memory for color alone, but the same amount as shape alone. Hence, the activity in the conjunction condition was as high as in the most difficult feature condition. A similar result was reported by Woodman and Vogel (2008) who analyzed the contralateral delay activity. Piekema et al. (2006) found activity in the right hippocampus for object-location binding but not in a feature memory task or for object-color binding. Location-specific (place-specific) activity was also found in a proportion of cells in the entorhinal cortex of monkeys (Suzuki et al., 1997). However, complete absence of any specific brain activity or even reduced activity in conjunction conditions was also reported. Conjunction tasks showed activities in the same posterior and parietal perceptual areas that were active in feature tasks (Elliot and Dolan, 1998; Sala and Courtney, 2007). However, whereas the activities were the same in the study of Elliot et al., Sala and Courtney reported even lower activities in the conjunction than in the feature condition.

Obviously, the empirical data regarding the neural basis of binding in WM are very heterogeneous. Differences between the tasks and the kind of to-be-bound features are probably the reason for this variability. Binding is not a unitary process. On the contrary, different types of binding exist that are based on different mechanisms and different neural structures (Zimmer et al., 2006a). There is evidence, for example, that feature binding within objects (intra item) is provided by different mediotemporal structures (perirhinal areas) than binding between objects (inter item) or between object and context. The latter two require the hippocampus (see Ecker et al., 2004). Moreover, the PFC plays a different role in these two binding mechanisms (Cabeza, 2006). Though this difference has not yet been considered in research on WM, I expect that it is relevant for these tasks, too, especially because location and object information should be bound by different processes (inter- versus intra-item, respectively). For example, in long-term memory tasks we could demonstrate that binding of object and color information has other electrophysiological signatures if color is an object feature than if it is a background feature (Ecker et al., 2007). Similar results are to be expected in WM tasks.

Another issue is different mechanisms of binding. One possible binding mechanism is provided by interneurons that connect to-be-bound assemblies. One might think on structures in the hippocampus and para-hippocampus (Knowlton and Squire, 2006; O’Reilly and Norman, 2002). A second mechanism is phase locking of neural activity between neurons that represent bound features (Raffone and Wolters, 2001; Vogel et al., 2001). Raffone and Wolters presented a computational model that simulates working memory as neural network. In this model, features are represented as activities of neurons in cell assemblies and neurons which code features of the same object oscillate in phase, i.e., they are synchronized. Empirical evidence for this synchronization was reported by Tallon-Baudry (2003). During a WM task, power in the induced gamma (24–60 Hz) and beta (15–20 Hz) band was
enhanced at frontal and occipital electrodes (Tallon-Baudry et al., 1998)—induced activities are caused by stimulus presentation but they are not time-locked to the stimulus onset as evoked activities. During maintenance, the occipital gamma and frontal beta component was most prominent and it was related to memory performance (Tallon-Baudry et al., 1999). Hence, this synchronization may be a mechanism of persisting feature binding in WM. Neural synchronization might also set the limit to WM capacity. In the model of Raffone and Wolters, the maximal number of items that can be stored results from the effect of inhibitory interneurons on assemblies coding different items. A sufficiently high activation of an assembly and temporal segregation of activities in different assemblies is only possible for a limited number of items (Murre et al., 2006). The efficiency of this network, e.g., the precision of synchronization, may even be responsible for interindividual differences. See Werkle-Bergner et al. (2006) for a thorough discussion of this possibility under the perspective of lifespan psychology and changes of neural processes in the course of aging.

2.5. Working memory as activated long-term memory

We begin to understand the neural basis of working memory and we see that these structures partially overlap with LTM. However, there are only few studies in which the neural structures that provide long-term and working memory are directly compared. Ranganath et al. (2004a) took a first step to disclose the relationship between neural processing in working memory and its consequences for long-term memory. The authors required their subjects to learn faces, houses, and face-house associations. The WM-paradigm was designed as a delayed matching to sample task (DMS), the LTM-paradigm as a delayed-paired associate task (DPA). In the latter, participants learned associations between a face and a house before the experiment, and at test, one element was given as a cue and the corresponding pair element had to be retrieved. In DMS, during encoding and maintenance, the fusiform face area was more active for faces than for houses, and inversely, the parahippocampal place area was more active for houses than for faces. In contrast, in DPA trials, this pattern was observed only at cue presentation, while afterwards the activation shifted to the domain-specific areas of the target. Additionally, in DPA trials, during cue presentation and probe presentation, the right anterior prefrontal cortex and the left hippocampus showed enhanced activity (for houses and faces). In another experiment, it was shown for novel objects that activity during WM maintenance in an S1–S2 task predicted subsequent long-term memory (Ranganath et al., 2005). Again, DLPFC (BA 9) and the hippocampus exhibited greater activity during maintenance for items that were later remembered than for those that were forgotten. In the late part of the delay period, occipital regions showed enhanced activity for remembered compared to forgotten items suggesting that these structures store the information. These data demonstrate that the storage sites of WM and LTM are very likely the same.

Bledowski et al. (2006) investigated retrieval from WM in a study that combined fMRI with ERP source analysis. Their participants encoded during study one or three sequentially presented nonsense shapes and after a delay of 7400 ms a target figure was shown to be judged as old or new. During encoding, strong activation in the inferior temporal and posterior parietal cortex was observed and the BOLD responses were enhanced for a memory set of three items compared to one item. These structures were also active during testing, but at that time they were not load dependent. In the test phase, load dependent activity was only observed at inferior frontal gyrus and at the right supplementary motor area (SMA). ERPs were identical for encoding and testing in the early time window. A transient activity in the inferior temporal cortex occurred that was independent of set size. It was interpreted as reflecting item encoding. This component was followed by a strong sustained activity in the posterior parietal cortex that did not vary with load although the posterior neural activity that generated these potentials did vary with memory load. Additionally, in the ventrolateral prefrontal cortex a transient activity was observed that was stronger for the larger item set, and this activity was interpreted as retrieval operation. These data are in good agreement with the idea of a fronto-parietal WM network and with the idea that WM is provided by the same structures that contribute to LTM.

Based on such data, Ranganath suggested a WM model that is fairly consistent with the view that has been developed in this overview so far (Ranganath, 2006; Ranganath and Blumenfeld, 2005; Ranganath and D’Esposito, 2005). Putting together their ideas and the pieces of information presented in this paper, the following picture arises (see Fig. 3).
Visual aspects of a stimulus are maintained in posterior visual cortical areas. This representation is hierarchical. At the lowest level, we find representations of sensory features (for visual input they are located in the occipital cortices), and at the highest level, objects are represented (in inferior temporal cortices). The more anterior the representations are located the less sensory specific they are. Representations of novel objects need medial-temporal structures (perirhinal cortex and hippocampus) as do conjunctions and relations between objects because new bindings between features have to be established. Posterior representations provide memory as passive stores. Ventral PFC (re)activates these posterior representations, and this facilitates maintenance. The dorsolateral PFC is relevant for the representation of complex stimuli and for manipulation (reconstruction, transformation, etc.). Spatial information (coordinates) is represented as retrospective code in the parietal cortex, and as prospective code in the superior frontal cortex. These spatial representations are supramodal. Additionally, spatial information can be provided as configuration of locations by the ventral system.

Active rehearsal is executed by the PFC operating on posterior storage structures. It provides a rehearsal network, may it be a prospective, a spatial or a visual code. This rehearsal process is supported by top-down input from medial-temporal structures especially after distraction (Passingham and Sakai, 2004). Kessler and Kiefer (2005) called this strategic process “endogenous recovery”, as opposed to “exogenous” recovery, that automatically occurs when the stimulus is re-presented. Reactivation is accompanied by shifts of attention to the processed information—or taking it the other way around, directing attention to a specific content reactivates this entry. WM of non-visual information is provided by the respective sensory processing structures, or in case of multi-modal representations, by neural structures receiving input from many modalities.

Most of these structures are likewise discussed as elements of a processing network relevant for LTM (e.g., Simons and Spiers, 2003). I therefore assume that both WM and LTM performances are provided by the same structures representing active items as synchronized activities in content-specific cell assemblies. However, both memory tasks have different characteristics and constraints because the demands of these two tasks are different. For example, in WM active but decaying neural structures have to be refreshed, whereas in LTM passive representations belonging to the same object have to be reinstated, and different cues are available for these tasks. Hence, WM and LTM may follow different memory laws even though the same structures store the information.

3. Imagery and VSTM

Until now we have mainly reported studies in which VSTM was investigated as means of short-term remembering. This is one function of VSTM and as Hollingworth et al. (2008) has shown it is also an important function of VSTM in everyday context. However, this is a rather passive task and because we are speaking of working and not only of short-term memory, we should also have a look at more active tasks. According to my understanding, VSTM is also involved in active visual processing. Prototypical active visual tasks are visual imagery, spatial rotation, and spatial reasoning. Consequently, imagery was used as main task in several behavioral studies (Logie, 1986; McConnell and Quinn, 1996; Quinn and McConnell, 1996). Likewise Baddeley (1976, 1986) considered VSTM crucial for image generation. However, current models of VSTM make different claims regarding this point. Whereas some authors (e.g., Quinn and McConnell, 2006) have suggested that the active component of VSTM is involved in imagery and visual perception, Logie (2003) has emphasized that the structures used for imagery are different from those used in VSTM. Generation and maintenance of images is seen as a function of the central executive, not of the visual cache.

This conclusion is partially based on results from studies using the mental synthesis paradigm, also called mental construction task (summarized in Logie, 2003; Pearson, 2001). Participants perform simple figures – usually presented as verbal descriptions – and they are instructed to combine these elements into an integrated and interpretable image. For example, an upper case “D”, “I”, and “V” are presented. When a participant mentally rotates the “D” 90° to the right, puts the “I” on top of it, and finally mounts the rotated “V” to the “mast”, an image results that looks somewhat like a sailing boat. According to Logie and Pearson, the active manipulation of the material should be performed by the central executive, the phonological loop should maintain the names of the symbols, and the inner scribe should temporarily hold the results of the manipulation. Taking this into account, Pearson (2001) proposed a revised model in which a visual buffer (similar to Kosslyn’s buffer) was introduced as an additional component used for the construction of a conscious image, under the control of the central executive. This buffer is the workspace for active transformations, whereas the visual cache is a back-up store for intermediate results. Which neural structures provide these components is not outlined.

Similarly, Quinn (2008) and Quinn and McConnell (2006) have postulated that imagery uses a visual medium, stressing however that this medium is shared with perception. This position is mainly based on results gained in interference studies. In these studies, participants perform an imagery task – usually the peg word mnemonic or the loci technique – and dynamic visual noise (DVN) is used as interference. DVN is a field of flickering white dots on a black background, which is presented as irrelevant visual input. It is assumed that this material involuntarily enters the visual buffer. The authors observed that DVN presented concurrently to image generation worsened memory performance. Because the visual complexity of DVN modifies the amount of interference (McConnell and Quinn, 2004), it is likely that the interference is visual. On the other hand, DVN was not effective when it was only presented during maintenance (Andrade et al., 2002; Avons and Sestieri, 2005; Quinn and McConnell, 2006; Zimmer and Speiser, 2002). Quinn therefore postulated a buffer component – similar to Kosslyn’s attention window – that is part of an image mechanism and that holds conscious images, see the discussion in Quinn (2008). Hyun and Luck (2007) observed a visual but not a spatial interference if during the retention interval a mental rotation task was performed. This suggests that a visual component as the assumed buffer may be the work place of mental rotation. Because the buffer is used by perception as well as by imagery and the perceptual input of DVN automatically enters the buffer, it interferes with imagery tasks. Independently of the buffer, a cache may hold information that is displaced from this conscious state. This passive information can be used to re-instantiate the content of the buffer when it was displaced from it.

Although the relationship between a visual cache and an imagery buffer is still controversial, two consequences of the discussion are already visible. First, a conscious image (e.g., the appearance of a star) is considered distinct from the representation of knowledge about its appearance, which allows for the generation of the conscious image. Second, visual images are considered elements of visual WM, represented in structures that are shared with perception. When visual input is displaced from conscious experience, it is not erased but it is still available if suitable retrieval cues exist. In Kosslyn’s model (1994), the former distinction is the one between an image held in the visual buffer and a compressed image. A compressed image is a feature vector.
having no topographic structure, but sufficient information to reconstruct an image. These two images are representations in two different structures, the buffer and the pattern activation system. However, one can also think about different states of representation within the same structure. It is difficult to distinguish between these two alternatives on the basis of memory experiments alone. Some neuroscientific data, however, suggest that the distinction is relevant. Uhl et al. (1990) reported slow potentials during image generation. Shortly after cue presentation, frontal and parietal occipito-temporal negativity was found, but only the posterior activity persisted. These areas were also found active in a PET study in which participants were explicitly required to use images to solve the task (Kosslyn et al., 1993). Imagery-specific activity was observed in the left DLPFC, left angular gyrus, right BA 19, right superior and inferior parietal lobe, left BA 17, and the fusiform gyrus. According to Kosslyn (1994) and Thompson and Kosslyn (2000), BA 17 implements parts of the visual buffer (maybe together with BA 18), middle temporal and fusiform gyrus represent and encode the pattern code, the posterior parietal lobes represent locations – metric coordinates in the right and categorical spatial relations in the left hemisphere – and the DLPFC is relevant for active processing (for specific details and for part processes of image processing see Kosslyn, 1994).

The mentioned structures are quite consistent with the structure that we had discussed as elements of the VSWM network. It is therefore likely that both tasks use the same network. Also the idea that levels of representations exist, from more specific to more abstract codes, was already discussed in the context of feature memory. It is therefore worth having a closer look at data from electrophysiological and fMRI studies on imagery. They will provide further support for the assumption that WM and imagery use the same network, but they also demonstrate that it is not a unitary structure.

3.1. Imagery of objects

In a number of image generation tasks, activation of the visual association cortex and additionally of category-specific brain areas was reported. D’Esposito et al. (1997) required their participants to generate images to words. They observed activity in the visual association cortex, specifically the left inferior temporal cortex (BA 37), extending superior into BA 19. Handy et al. (2004) compared recall of pictures and image generation and found fairly overlapping brain activity. In the frontal cortex image generation was somewhat later than picture retrieval, and both tasks activated the same left ventral temporal cortex area, including the posterior fusiform cortex. In a similar imagery task, inferior temporal activation was also reported by Mellet et al. (1998). Kosslyn et al. (1999) presented their participants meaningless visual patterns, which they had to impose in order to ‘read out’ specific details they were asked for some time after stimulus presentation. The authors observed activity in BA 17, and TMS in this area increased reaction times.

Domain-specific activation was found when tasks were presented making reference to the specific category information. O’Craven and Kanwisher (2000) reported activity in the fusiform face area and parahippocampal place area, when faces or places were imagined, respectively. Perceptual-specific areas were activated when participants imagined faces, houses or chairs (Chao et al., 1999; Ishai et al., 1999; Mechelli et al., 2004). Ishai et al. (2000), also reported content-specific activity in the temporal lobe, and they pointed to the fact that activity in the parietal and frontal cortex was not content related. Ishai et al. (2002) found that a subset of regions in the ventral path that contribute to perception of faces was also active during imagery. However, imagery showed additional activity in the IPS and precuneus which increased if details were relevant. Goebel et al. (1998) found activity in area MT for the imagination of movements. Kourtzi and Kanwisher reported activity in MT/MST if participants viewed static photographs with implied motion. Halpern and Zatorre (1999) investigated auditory images (melodies) and reported activities in associative auditory cortex (BA 21/22). Culham et al. (1998) asked participants mentally to track bouncing balls and they observed activity in the lateral parietal cortex and the precuneus.

Mental navigation on a known route activated the precuneus, the insula and the hippocampus (Gaem et al., 1997). Motor imagery activated the SMA and premotor cortex (Dechent et al., 2004) and also the superior temporal sulcus (see Decety and Grèzes (1999) and Jeannerod (2001) for review).

Summarizing these studies, we can conclude that during imagery brain activity was consistently observed in those areas that process the specific type of information during perception or action. Motor information is located in the frontal cortex, whereas visual information is located in the temporal and tempo-occipital cortex. A controversial topic is whether in visual imagery activity is found only in the association cortex or even in primary sensory cortices (e.g., BA 17). In a meta-analysis of imagery studies Kosslyn and Thompson (2003; Thompson and Kosslyn, 2000) tried to isolate methodological characteristics of studies that are critical for the activation of BA 17. They concluded that the early visual cortex is only activated if a detailed visual image was generated. However, not all agree on that which is why this topic remains controversial. Nevertheless, independent of this controversy, the results speak in favor of the assumption that imagery of sensory information activates brain areas in the occipital and temporal lobes that correspond to the category and modality of the target information. Additionally, they suggest that within the occipito-temporal cortex along the anterior–posterior axis a concreteness gradient exists with very specific representations at the posterior and more abstract representations at the anterior pole.

3.2. Spatial imagery tasks

In contrast to the former studies, activity in parietal cortex, specifically in the IPS, was found when spatial imagery tasks were required. In several of these studies maps were presented. The authors required their participants to inspect a mental image of a studied map (mental exploration) (Mellet et al., 1995), to scan along a path on a mental image of a map (mental scanning) learned from either inspection (Mellet et al., 2000a,b) or verbal descriptions (Mellet et al., 2002), or to visualize scenes studied before (Mellet et al., 2002). In all these studies, activity in the IPS was found (see Mazard et al. (2004) for a review). Trojano et al. (2002) investigated mental activity in a clock task. They asked their participants to imagine an analogue clock showing acoustically presented times, and to judge which of two clocks had the larger angle between its hands. In this task, the IPS was specifically active (see also Trojano et al., 2004). Sack and colleagues (Formisano et al., 2002; Sack et al., 2002) stimulated the parietal cortex with repetitive TMS (rTMS) during task processing, and they observed a selective enhancement of the fMRI signal in the IPS without rTMS but a selective impairment in this structure after rTMS. Obviously, different from imagery of objects, spatial imagery activates the parietal lobe.

Area IPS was found active also in a number of transformation tasks. A typical task in which the active component of WM is needed is the transformation of visual images. An example is the mental rotation task of Shepard and Metzler (1971). Participants are to mentally rotate a reference figure in order to decide whether it matches the target presented in a different orientation. Heil (2002) reviewed a number of studies that measured slow
potentials during mental rotation. In general, a negative potential at parietal electrodes was observed (Heil et al., 1996) that increased with angular disparity (Rösler et al., 1995). Interestingly, the IPS was also active in a tactile version of the task (Röder et al., 1997). This result shows that the parietal activation is not a consequence of visual input, but it is probably a correlate of spatial processing itself. Data from an fMRI experiment support this conclusion (Jordan et al., 2001). The superior and inferior parietal lobe (around the IPS) was active during mental rotation with three different types of input material (letters, 3D figures, and abstract figures). Similar activations were also reported for 2D and 3D figures (Suchan et al., 2006). Cohen et al. (1996) contrasted a rotation condition with a simple comparison of the two figures without rotation, and they reported enhanced activation in the parietal cortex (BA 7), in midfrontal areas (BA 8), and also in parieto-occipital areas (BA 39, 19, 18).

Activity in superior parietal lobes was also observed by Richter et al. (2000), but they found activity additionally in premotor areas. The observation that brain structures related to motor execution are active in a rotation task is plausible if participants solved the task by internal action, i.e., turning the figure by imagined hand movements. Kosslyn et al. (2001) demonstrated that this difference is in fact critical. When participants were trained to solve the task by imagined actions, a strong activation of the (pre)motor area was observed compared to an untrained condition. Similar results were reported by Vingerhoets et al. (2002). They emphasized that only left motor areas were active (those of the dominant hand) when tools were shown, but bilateral activation was found when pictures of hands were shown. This is compatible with the observation that imagination of overt actions (mentally simulating performance) activates premotor areas (see Jeannerod, 2001). A direct relevance of the motor area for motor imagery could be demonstrated by Tomasoni et al. (2005). They stimulated a grid array placed directly over the left primary motor area in a patient. They observed an interference effect on a motor imagery task but not on a visual imagery task. Activities in motor areas during mental rotation were also found in other studies. Usually the SMA was found to be active. Windischberger et al. (2002) observed activity in the parietal cortex and in the mesial part of the SMA, and also Johnston et al. (2004) reported activity in the SMA, in the parietal cortex and in the ventrolateral PFC. A variant of mental rotation that simulates external actions obviously activates areas that are involved in the control of overt motor actions.

However, in spatial transformation tasks without a motor component, usually parietal structures and especially the IPS were active and this activity was clearly related to task performances. Tagaris et al. (1996, 1997, 1998) found a positive correlation between error rates and parietal activity whereas right precentral activity was correlated with the speed of mental rotation. Carpenter et al. (1999) reported a linear increase of activity in IPS with the amount of angular rotation. Just et al. (2001) varied the complexity of the task by manipulating the rotational axis and the number of steps needed to solve the task (imagine an old fashioned alarm clock from different viewpoints) and the way of rotation was defined by verbal commands. They observed an increase of activity with task complexity in the parietal cortex, but also in the left inferior frontal cortex and bilateral DLPMC. Compared to the rotation of perceived stimuli (as in the Shepard mental rotation task presenting two stimuli side by side), the parietal effect was similar but the extrastriate cortex was much less activated than with real stimuli. Therefore, though both structures were found active in mental rotation tasks, they may perform different part processes. Koshino et al. (2005) simultaneously varied angular rotation and visual complexity and they reported interactive effects. Activities in ventral and parietal posterior areas as well as in prefrontal areas varied with task difficulty, but the effects were not additive. Neural activity in the parietal cortex seems to be confined to explicit object rotation. Only explicit rotation changed parietal activity, whereas presenting objects from different viewpoints for recognition modulated activity in the ventral path (Gauthier et al., 2002). Direct evidence for a causal role of the parietal cortex in mental rotation was found in rTMS studies. Harris and Miniussi (2003) reported interference on mental rotation after a stimulation of the right parietal cortex, which was effective, however, only in a small temporal window after stimulus presentation (400–600 ms). Zacks et al. (2003a) applied rTMS to the parietal cortex and this impaired mental rotation of objects, but not rotation of one’s own body in a scene (perspective taking). This is compatible with the observation that different neural structures provide allo- and egocentric rotations. Object rotation increased activity in the right parietal cortex and decreased activity in the left, whereas perspective transformations activated the left temporal cortex (Zacks et al., 2003b).

These data highlight the relevance of parietal structures for mental rotation. The parietal cortex provides mental object rotation, but probably not rotation of one’s own body. However, the data also illustrate two unsolved issues in the interpretation of relationships between neural activities and task performances. One is the variability of strategies between participants and tasks (see, e.g., Butler et al., 2006). Most cognitive tasks can be solved by different means and the activated brain structures very likely co- vary with the type of strategy used. Related to this problem is the fact that each task consists of several part processes which are very likely provided by different brain structures. In order to make further progress in the neurocognitive understanding of imagery processes we need computational models. We have to specify the cognitive components and their neural realization of the different (part) processes that enable the generation and use of images (see, e.g., Kosslyn, 1994). Second, these strategies and the brain activities interact with participants’ abilities. Inferences about part processes often make use of a parametric manipulation of task difficulty and co-varying changes of neural activity are observed. However, it is all but clear what higher or additional neural activations mean in different brains mean. Better performing participants may be more talented and they may need less neural activity because their neural structures are more efficient (Gray and Thompson, 2004). Alternatively, they might be better because they use more neural substrate to solve the task because they work harder. Vitouch et al. (1997) observed slow negative potentials in a visuo-spatial imagery task (3D cube test) and the amplitude changes of these potentials negatively correlated with participants’ abilities. This is compatible with the “neural efficiency” hypothesis. Lamm et al. (2005 #6623) did a similar study, but found that the activation for high and low performers did not differ and the neural network that was found active was the same. This might be the consequence of a tradeoff between efficiency and effort. Heitz et al. (2008) investigated WM capacity by pupillometric methods and the authors actively varied effort by incentives. Incentives increased capacity and effort in terms of pupil size. Because it was an additive effect it is likely that without incentives high and low performers applied the same effort. Considering these strategic influences, we probably need multi-method paradigms to reach conclusive results about the relationship between individual imagery abilities, VSWM capacity, and the neural mechanisms that cause these efficiencies and their limitations.

3.3. Reasoning and VSWM

As in mental rotation, imagery processes are involved in a number of other cognitive tasks. Examples are the generation of
mental models (Glenberg and McDaniel, 1992), comprehension of descriptive texts (e.g., a description of a spatial environment, Denis and Zimmer, 1992), spatial reasoning (Fiore and Schooler, 2002), or mental animation in mechanical tasks (e.g., verifying whether moving a lever in a given direction will move a connected mechanic component into a specified direction, Hegarty and Just, 1993). Also studies using these paradigms brought evidence in favor of the assumption that VSWM and imagery is provided by the same system. For example, when sentences were presented that expressed a visuo-spatial relation (e.g., A tractor has larger wheels in the rear than in the front.) their verification took longer in written than in oral form (Eddy and Glass, 1981; Glass et al., 1985).

Similarly, during comprehension of a descriptive text that was presented together with a picture of the issue, a visual memory load interfered with comprehension, whereas an oral preload did not (Krueley et al., 1994). DeBeni and colleagues systematically compared different forms of presenting instructional texts with different demands on visuo-spatial processing (De Beni and Moe, 2003; De Beni et al., 1997). An oral presentation caused a better performance than a written one in an imagery task, and vice versa for rehearsal. In a later study, these two main tasks were combined with a verbal (articulatory suppression) or a spatial (spatial tapping) secondary task. Articulatory suppression interfered with the verbal and with the spatial task, whereas spatial tapping only interfered with the spatial main task (De Beni et al., 2005). Finally, in mechanical reasoning tasks (mental animation) participants with high visuo-spatial abilities outperformed those with low abilities (Hegarty and Sims, 1994), and again, a visuo-spatial load impaired performances (Sims and Hegarty, 1997). These results are plausible if imagery and visual perception use the same part system which has a limited capacity and because neural structures of perception and VSWM overlap imagery and VSWM share neural structures.

Some of these reasoning tasks were also investigated by neurocognitive methods and these studies allow a direct test of this hypothesis. Just and colleagues investigated brain activations during verification of high imagery sentences. For high imagery sentences, the authors observed activity specifically in the left intraparietal sulcus (Just et al., 2004). Explicit imagery of geometric figures activated the visual association cortex (BA 19) and the parietal cortex (BA 7 and 40) (Knauff et al., 2000). Knauff and colleagues also investigated the construction of mental models in spatial reasoning tasks. They again found activation in the superior and inferior parietal cortex, the precuneus, and the visual association area together with activity in the prefrontal areas (BA 6 and 9) (Knauff et al., 2002). Later they contrasted conditions in which a more abstract mental model was sufficient with conditions in which the generation of a more concrete mental image was required. In this study they could demonstrate that the superior parietal cortex was active in both tasks whereas in the imagery condition the visual association cortex (V2) was active in addition (Knauff et al., 2003). When maintenance and reasoning were separated, the posterior areas were more active during maintenance and the prefrontal areas during reasoning (Ruff et al., 2003). All these structures have already been found active in VSWM tasks. Taken these results together, I come to the conclusion that VSWM and visual imagery are partially provided by the same neural network.

This conclusion, however, leads to some problematic issues because imagery and WM tasks did not always show the same results. Quinn (2008) therefore made a distinction between a visual cache which passively holds sensory information and the visual buffer which is a mental workspace for imagery. On the other hand, in the review many studies were reported showing that early occipital and temporal areas were active in both tasks. What is a plausible solution for this contradiction? In my view it will be found in the distributed and hierarchical character of the representation. The same temporo-occipital and parietal network may be in different states dependent on the task. The task configures the anterior–posterior circuit. This also changes consciousness of the experience/memory. At the behavioral level, one might think on attention and at the neural level on synchronized activity between different brain regions. Further research in the time-frequency domain of neural activity in WM and imagery tasks is necessary here.

4. Concluding remarks

Fig. 4 summarizes the results of the imagery studies and it maps these data onto the structures we had previously discussed as

![Image: An illustration of different imagery components and the neural structures that might realize these functions during imagery and working memory.](image-url)
constituents of the WM network. As one can see, similar structures were found active in imagery and visuo-spatial reasoning tasks as I previously discussed in the context of in perception, and this network is also active in perception (see Farah, 1989). The dorsal-ventral distinction was relevant also in imagery tasks and prefrontal activities were observed varying with task complexity. Additionally, again, a network of brain structures and not a single unit is involved in image generation and processing. Furthermore, domain-specific brain areas including motor areas belong to this network.

Inferior temporal and temporo-occipital brain areas in the ventral path were active in a similar way during imagining objects and during remembering the visual appearance of objects. Parietal structures were active in spatial imagery tasks and these structures were also active in spatial WM tasks. In the context of WM, we had discussed that this part of the brain may be a modality-unspecific storage component of location information. The IPS is involved in mental rotation together with premotor and supplementary motor areas. Remember, that the IPS is involved in processing the orientations of 3D objects for manipulation. The demands of this task are very similar to the demands of mentally rotating objects. The inferior part of the parietal cortex seems to process more detailed spatial models. This activity often extends into the precuneus—an area that was found active in WM tasks, too. The precuneus seems to be involved in many imagery and mental simulation tasks (see Cavanna and Trimble (2006) for an extended review). Additionally, activity in the precuneus was also observed in LTM tasks after imaginary encoding (e.g., Fletcher et al., 1996; Lundstrom et al., 2003, 2005; Shallice et al., 1994). For this reason, Fletcher et al. called this structure the ‘minds eye’. It is possible that this part of the brain provides a workplace for processing the visual appearance of perceived and generated images in the narrower sense. It is also possible that this specific structure is a critical part of the anterior-posterior circuit that I mentioned above and in order to experience a conscious image. However, WM is not restricted to this area. The intra parietal sulcus seems to have a specific importance for active visual imagery and processing, especially when spatial information has to be transformed and when images have to be manipulated actively. Motor areas were found active in motor imagery tasks and in reasoning about motor acts. Superior temporal sulcus and MT were involved in imagery of biological and object movements.

From the perspective of the involved neural structures, we should therefore not speak of a specific subsystem that is dedicated to WM. We never found activity in an isolated circumscribed structure. It was always a network of neural structures that provided the different performances and this network was assembled in different ways according to demands of the task. Instead of a dedicated working memory system, it therefore seems to be more adequate to speak of different tasks with specific demands. These tasks are associated with mental operations running on representations each realized by specific neural structures and a part of them is domain-specific. One task is temporary storage of specific information of a stimulus that no longer exists in the outer world (passive WM); another task is retrieving information that was encoded some time ago (restate-ment of LTM content in WM); a third task is imagining contents retrieved from semantic memory (image generation in WM); a fourth task is transforming perceived stimuli (active WM), and a fifth task is reasoning (using WM contents in problem solving). Performances in these tasks differ because the tasks have different demands. For example, WM and LTM tasks use different retrieval cues. Cues for WM were always encoded recently (a few seconds ago), whereas LTM uses cues that were encoded more distantly in the past. Therefore, LTM rely more strongly on consolidation processes than WM. Furthermore, although the same neural structures provide long-term and working memory, they probably provide them by different neural mechanisms. Entries in WM may be represented as synchronized cell assemblies whereas these networks have to be consolidated in order to be available as entries in LTM (Meeter and Murre, 2005; Paller, 2000). Hence, even if WM is provided by the same neural structures as LTM this does not imply that both memories work identically and performances depend on the same variables. Even if the same structures contribute to both memory tasks, they may be embedded in different networks, so that they can behave differently.

In closing, one might wonder whether VSWM now exists or not. It does so, if we take it as an umbrella term for a family of cognitive tasks and their neural realization. It does also, if we refer to neural networks that make representations temporarily available. However, it does not exist in the sense of an independent separate structure existing besides perception and long-term memory. Even though WM is a specific memory task it is not necessary to postulate additional memory buffers. It is sufficient to postulate additional processes and to allow that the same neural structures can take part in different networks that are task-specifically configured. In this view, WM is a by-product of processes running on neural representations that are used in specific tasks including perception and reasoning. Recently, Postle (2006) made this hypothesis explicit.

The general logic is the following. Brain structures represent parts of the environment during perception and action. Some of them are modality-specific others are supra-modal. As long as these structures are active, the information is (consciously) available. When the external activation stops (passive storage), representations start to decay and therefore shortly after removal of a stimulus its representation has to be re-represented either by an endogenous or an exogenous trigger (Kessler and Kiefer, 2005). Therefore, representations remain active as long as they take part in mental process, and during that time we may consider them as being in WM. A simple process that has this consequence is directing attention to a stimulus, but we can think of many other processes like mental rotation, mental scanning, synthesis tasks, mental animation, etc. Hence, temporary WM is probably an emergent property of any cognitive process. In the case of visual working memory it is any information provided by visual perception and processes running on such representations.

Acknowledgments

This research was supported by a grant of the Deutsche Forschungsgemeinschaft within the Collaborative Research Centre for Resource-Adaptive Cognitive Processes (SFB 378). I would like to thank the editors, two anonymous reviewers, Gerry Quinn and Katja Umla-Runge for helpful comments on an earlier version of this paper.

References


