Utrecht University

HUMBOLDT-UNIVERSITÄT ZU BERLIN

The Role of Networks in the Brain

in the Control of Remembering and Forgetting Working Memory Items

Laura van Zantwijk (6318290)*

MSc Neuroscience and Cognition Minor Internship Thesis

Supervisor: dr. Thomas Christophel

Examiner: dr. Surya Gayet

Abstract

Working memory items can be quickly stored, updated, and removed. In the current study, we aimed to identify what neural mechanisms are involved in the executive memory control of keeping an item in working memory or removing it. Additionally, we aimed to determine if a potential distinction between remembering and removing is consistent across different sensory modalities, that is across visual and auditory information. We reanalysed the data of a retro-cue task and used multivariate pattern analyses to identify the patterns of neural activity that realise the remembering and removal of working memory items. We did so by training on both visual and auditory sensory modalities and decoding the timing of the cue (early or late). The timing of the cue reflected the instruction to remove an item from working memory or keep remembering it. We found that several networks play a role in controlling the remembering and removal of working memory items during the delay period of the retro-cue task. These networks include the dorsal and ventral frontoparietal attention networks as well as the default mode network. Importantly, these findings could not be explained by the modality of both the cue and the sample stimuli. Results suggest that the control over the decision to keep an item in working memory or remove it may not be localized to specific areas in the human brain. Instead, the cognitive process of updating the items that are in working memory might be so big that it is distributed throughout numerous cortical areas.

Keywords: working memory control; forgetting; multivariate pattern analysis; CvCrossManova; attention networks

Layman Summary

This study looks at which parts of our brain are involved in keeping and forgetting information that is in our working memory. Working memory is like a notepad in our head that helps us to hold and work with information for a short time. We wanted to see if this process of making such a decision in the brain is the same for different types of information. These types of information can be things we see (visual information, like a coloured object) or hear (auditory information, like a music note). People took part in a task where they had to remember certain things and forget other things based on cues given to them. A cue is a signal that tells people which things to focus on and which to ignore. They did this task while they were laying in an fMRI scanner. fMRI stands for "functional magnetic resonance imaging". An fMRI scanner measures brain activity by making a scan of the brain. It is a special kind of scan that shows which parts of the brain are active when we do different things or think different thoughts. It does so by measuring changes in blood flow of the brain, which can indicate when certain parts of the brain are working harder. We then looked at these brain scans using multivariate pattern analysis. This is a common technique in fMRI where we can look at patterns of activation and predict multiple results that might come out of it, called "decoding". What this means is that we take the activity patterns from the brain scan and show it to a computer. We use these activity patterns to teach the computer what brain activity looks like while people are performing our task in the scanner. After the computer is trained, it can look at brain scans it has not seen before and tries to guess ("predict") what the person was thinking or doing. Using this method, we found that several brain networks help us to control what stays in and what gets removed from our memory during the task. These brain networks include two networks that are involved in attention, and a network usually active when we are at rest. Importantly, this control process seems to involve combinations ("networks") of multiple brain regions, rather than any individual brain region. Additionally, it does not depend on whether the

information is visual or auditory. This means that updating our working memory might be such a complex process that it affects many parts of the brain.

Introduction

In the human brain, sensory working memory plays a crucial role in temporarily holding on to information about stimuli even after they are gone (Baddeley, 1992). However, our working memory has limits, meaning only a fraction of the vast sensory representations can be kept at any given time (e.g., Constantinidis & Klingberg, 2016; Cowan, 2010; Ma et al., 2014). Additionally, there can be interference between the sensory representations we are trying to remember (e.g., Jonides & Nee, 2006; Oberauer et al., 2012), complicating this process further. Given the continuous input of sensory information, managing the process of holding on to information becomes complex.

To deal with this complexity, we have a way of deciding what is important and what can be ignored, so that sensory representations that are deemed irrelevant are not retained further. This selection of the to-be-memorized contents is by some believed to be governed by a common memory control mechanism, often considered to be part of the 'central executive' of our memory system (e.g., Baddeley, 1992; Collette & Van der Linden, 2002; Funahashi, 2017; Wongupparaj et al., 2015). Therefore, to control memory is to harbour the information of which sensory representations should be remembered. This also involves strengthening those selected working memory representations or weakening those that are deselected (Christophel et al., 2024).

We could wonder what working memory representations even look like in the brain. Early working memory models highlight the importance of the existence of storage buffers, which propose dedicated sites for the storage of different types of sensory information (Baddeley, 1992). For example, tasks tapping into phonological storage typically show activation in language areas such as Broca's and Wernicke's, whereas visuospatial tasks activate posterior regions like the parietal cortex (Smith & Jonides, 1997). Alternative models of working memory are the state-based models, which argue against such dedicated storage

sites. Instead, they propose that working memory relies on temporarily activating representations from long-term memory by directing your attention to them (Cowan, 2005; Oberauer, 2001), or by recruiting perceptual and motor systems (Postle, 2006; d'Esposito, 2007). This way, information can either be 'in' or 'out' of working memory by strengthening or weakening activation for it. These models do not yet provide clarity on whether working memory and long-term memory are then colocalized - suggesting that working memory activation is simply activation of representations in long-term memory - or if these memory representations reside in distinct areas in the brain. A third way of approaching working memory storage is through storage *networks*. Such networks are typically hypothesized to contain distributed patterns of activity for different memory representations (e.g., Christophel et al., 2017; Petersson et al., 2006). Cognitive processing then emerges because of the interaction between these brain regions (e.g., Petersson et al., 2006; Arbib, 2003; Mesulam, 1998). Overall, these three approaches – storage buffers, state-based models, and distributed storage networks – provide different frameworks for understanding how working memory representations are organized and maintained in the brain.

Throughout the years, various distinct brain areas have been the focus of investigation for storing working memory contents. The most prominently researched areas include prefrontal (e.g., Curtis & D'Esposito, 2003; Goldman-Rakic, 1995; Riley & Constantinidis, 2016; Mendoza-Halliday, 2015; Narayanan et al., 2005), parietal (Marshuetz et al., 2000; Xu & Jeong, 2015), and sensory (Awh & Jonides, 2001; Emrich, 2015; Harrison & Tong, 2009; Pasternak & Spinelli, 2015; Serences et al., 2016; Sreenivasan et al., 2014) cortices. For example, in an early non-human primate study (Jacobsen, 1936), lesions were made to the monkey's prefrontal cortex (PFC). They found deficits in the ability to maintain important information, while performance remained intact for tasks that did not require memory. This finding led most early electrophysiological studies to focus on the PFC, but later recordings in

monkeys have also revealed persistent activity for working memory representations in different areas. More specifically, research has identified specific brain regions associated with different types of visual processing: V1 for processing spatial location (Supèr et al., 2001), MT (Bisley et al., 2004) and MST (Mendoza-Halliday et al., 2014) for motion, V4 and the temporal cortex (Miyashita & Chang, 1988; Miller et al., 1993; Chelazzi et al., 2001) for shapes and real-life stimuli, and in the inferior temporal cortex (Fuster & Jervey, 1981; Fuster & Jervey, 1982) for processing colour information. In humans, identifying the brain areas that are involved in content-specific working memory is challenging with non-invasive imaging techniques, due to their relatively low spatial resolution (Christophel et al., 2017). However, using multivariate pattern analyses (MVPA), different types of stimuli can be read out from different areas of the human brain. For example, low-level visual features such as colour, orientation, motion, or more complex patterns can be decoded from early visual areas (Harrison & Tong, 2009, Serences et al., 2009; Pratte & Tong, 2014). Auditory stimuli can be read out from the auditory cortex (Linke et al., 2011; Kumar et al., 2016), and more complex shapes (Christophel et al., 2013) and orientations (Ester et al., 2015) from the frontal eye fields (FEF). In conclusion, extensive research has explored various regions for their roles in storing working memory contents, with recent advancements in multivariate pattern analyses further enhancing our understanding of this process in humans.

For working memory contents to be able to stay in working memory, they must be actively maintained for ongoing cognitive processing (Oberauer, 2009). This strengthening of working memory items to remember them is important, and its crucial counterpart is the weakening of items to remove or overwrite them. Removal clears the space for currently relevant input, addressing the overwhelming amount of information we are confronted with (Dames & Oberauer, 2022). The interplay of these two processes is essential for a comprehensive understanding of working memory dynamics and control: as we expect

information to become irrelevant over time, updating our working memory requires us to swiftly discard irrelevant contents to prevent clutter (Björk, 1970; Hasher et al., 1999; Oberauer, 2018). It has even been suggested that removal from working memory is an active updating process, akin to remembering (Ecker et al., 2014). Thus, the ability to remember and remove working memory contents is fundamental for effective cognitive processing and the overall control of working memory.

Behaviorally, research has investigated removal using retro-cue paradigms. In these tasks, participants typically encode a set of items into working memory, and then are cued to remember one of them (Souza & Oberauer, 2016). After a delay, a recognition probe is presented, and participants are to decide whether it matches an item in the relevant subset, thereby rejecting probes matching an item in the irrelevant set. Evidence for removal comes from the finding that, after the retro-cue, participants find it easier to add further information to working memory. This suggests that the previously taken spots in working memory have now freed up (Gunseli et al., 2015). Additionally, when memory for an irrelevant item is tested after a delay, it is extremely poor. Accordingly, representations of irrelevant information diminished to a nonsignificant level after a cue-stimulus interval of 1 second or more (Oberauer, 2001). Intrusion effects, meaning the mixing up of memories that are similar, were observed for all cue-stimulus intervals up to 5 seconds. This suggests that the focus of attention indeed weakens irrelevant representations. Together, this shows that the dynamics of how the brain deals with irrelevant representations might mirror the efficiency with which people can focus their attention on one of the two sets in working memory. In another retro-cue paradigm, a cost for irrelevant information was illustrated by a decrease in recall performance for information participants were pre-informed would be irrelevant, yet were unexpectedly tested on in a subset of the trials (van Moorselaar et al., 2015). The degradation of irrelevant representations persisted even when memory was loaded with only two items, well within the

typical capacity of working memory. All in all, retro-cue paradigms can be used to study the removal of irrelevant representations from working memory, clearing the space for new representations.

Similar evidence for removal exists on the neural level. Recent electrophysiological work in non-human primates has revealed that neurons in prefrontal and posterior areas show neural activity that reflects the identity of a cue. (Panichello & Buschman, 2021). Such cues instructed the animal which of the two items to remember. This activity, specific to the selected item, is important for any process exerting control over memory storage because it provides information about what needs to be remembered later. Initially, these findings were seen as evidence that the lateral prefrontal cortex acts as a domain-general controller. However, the study also observed selection-specific activity in the same brain areas responsible for memorizing contents (i.e., that carried activity specific to the content), and these two effects were not independent of each other. Similarly, in different prior work selection-specific activity was confounded with the perceptual representation of the cue stimulus (Quentin et al., 2019; Zhou et al., 2022). This makes it hard to conclude what process was driving the effect. Consequently, although these studies led to some useful insights about the control of working memory, in these tasks it is impossible to separate control from storage (Christophel et al., 2024).

The current study presents a reanalysis of the data from the retro-cue task of Christophel et al. (2024), since they were able to separate the storage and control as well as the remembering and removal of working memory representations. In the study, participants had to memorize two sample stimuli that were presented after another. This included a Gabor patch presented visually, and a pure tone presented auditorily. Participants had to retain both items until an informative cue told them which of the two had to be used for the task that came right after (see Figure 1). These cues were presented either early (8.575 s) or late (18.375 s) in the delay

and presented either visually or auditorily. Additionally, various representations of cue stimuli were used to instruct the selection of memory items. Using the data from their task, we can look at "selection-specific activity", that is activity related to selecting an item for further processing, which we define as the "control" of working memory. The timing of their cues, either early or late, reflect the instruction to store ("keep") or remove ("drop") representations from working memory. The rationale behind this is that an early informative cue in the trial indirectly instructs participants to drop the sample that is not cued, whereas participants had to keep remembering both samples throughout the delay in case the cue was late. More details on this retro-cue task and the analysis will be explained later in the current paper.

In the field of memory control, substantial focus has been placed on attempting to decipher complex phenomena. Think for example of mental control failures of reading a page in a book only to realize, upon reaching the bottom, that you have no recollection of the content (Hasher et al., 1999). Although such phenomena are undoubtedly interesting, can we step back and study memory control at the simplest level we can think of: What neural mechanisms are involved in controlling whether an item stays in our working memory or gets removed? We hypothesize that the control of remembering and forgetting different types of sensory information may rely upon a shared neural substrate. Multivariate decoding of different spatial locations and cue types has been shown to generalize across different sensory modalities (e.g., Antono et al., 2023; Christophel et al., 2024; Wu et al., 2018). Therefore, we hypothesize that the control mechanisms for different types of sensory information will also generalize.

To avoid selecting regions of interest (ROIs) and thereby prioritize certain brain areas over others, we opted for a searchlight based decoding approach. This way, we were less likely to miss out on brain areas that might be important. However, we can still hypothesize about candidate regions for involvement in exerting control over remembering and removal processes in working memory. Four regions that are known to be involved in selection of working

memory items are the lateral prefrontal cortex (lPFC), frontal eye fields (FEF), parietal cortex, and intermediate visual area V4, as has become evident from various fMRI studies in humans (see for example Nobre et al., 2004; Nee & Jonides, 2009). This is consistent with recent findings in non-human primates (Panichello & Buschman, 2021), where it was also found that neurons in all four regions carried information about which item was selected from working memory. Similarly, when visual cues were used, spatial representations in visual, parietal and frontal cortex have shown selection-specific activity and may therefore be recruited for control (Quentin et al., 2019; Zhou et al., 2022). Using searchlight based MVPA, it was recently demonstrated that the left middle temporal gyrus (MTG) and right intraparietal sulcus (IPS) contain representations of cue instructions that are necessary to implement working memory control (Christophel et al., 2024). These areas could also underly the control of the remembering and removal of sensory information, and we might therefore see distributed patterns across the brain. MTG is known to be involved in processing abstract, semantic information, and temporal order (Davey et al., 2016), whereas IPS is known for its involvement in spatial processing and attention (Gillebert et al., 2011). Potentially, the neural mechanisms of memory control are similar to processing abstract information (area MTG), and spatial processing and attention (area IPS). As we can think of memory control as an abstract process that requires attention, it would make sense to find that these areas harbour representations information. Finally, the *storage* of working memory contents has been shown to involve multiple brain areas (e.g., Lara & Wallis, 2014; Fuster, 1995) and is distributed across the brain (Christophel et al., 2017). So perhaps, in line of the idea of storage networks, working memory control of remembering and forgetting items is also not restricted to just a few areas. Instead, it might rely upon network activation. All in all, the current study is exploratory in nature, hence there are numerous different candidate areas we could find.

The current study presents multivariate pattern analyses performed on the dataset from Christophel et al. (2024). By looking at "selection-specific activity" – defined as activity related to selecting an item for further processing - we aimed to study the "control" of storing or removing working memory items. We did so by contrasting conditions where the timing of the cue was early ("drop" an item from working memory) or late ("keep" both items). Since the task was designed in such a way that both stimuli and cues were presented both visually and auditorily, we could also investigate whether indeed this control mechanism generalizes across different types of sensory information.

Methods and Materials

The current study presents a reanalysis of data from Christophel et al. (2024). Therefore, the methods and materials of these two papers are nearly identical.

The dataset is particularly suitable for our purpose of studying the control of remembering and forgetting. In the retro-cue study, memory control and storage were separated in three different ways. First, by using temporal-order cues, activity related to selecting an item ("selection-specific" activity, executive control) could be separated from activity related to storing the item ("content-specific" activity). Second, selection was separated from storage for both visual and auditory stimuli so that changes in activity for both stimulus modalities can be investigated. This resulted in separable brain regions for stimulus decoding. Finally, the response evoked by the cue was controlled for by generalizing across auditory and visual information. This allowed for the testing of generalization across these different types of sensory modalities. The way the task is designed, selection-specific activity should not be confounded with a change in content-specific activity. Importantly, the timing of the cues, either early or late, indirectly reflect the instruction to keep remembering or to remove representations from working memory. This task design is therefore highly suitable for the test of which parts of the brain carry information that is involved in exerting memory control over remembering and forgetting items.

Participants

The study contained thirty-six right-handed participants (20 female; mean age: 25.7 years ± 0.52 SEM) with no history of neurological or psychiatric disorders who gave informed consent. The authors determined the sample size using exploratory analyses in their prior work. The local ethics committee gave approval of the study.

Experimental Procedure

The task was an audio-visual retro-cue task (see Figure 1). Each trial, participants were presented with two sample stimuli consecutively, with a counterbalanced presentation order. These samples included a visual stimulus and an auditory stimulus (0.8s each, 0.2 s ISI). The visual stimulus was a sinewave grating of varying orientation (8 different orientation angles). The auditory stimulus was a pure tone of varying pitch (8 different pitches). Participants were instructed to retain both items (0.5 s) until a cue indicated whether the first ("1" or "A") or the second ("2" or "B") item had to be used for the task that came after. These cues were informative and were presented either auditorily or visually and either early (8.575 s after the first stimulus onset) or late (after 18.375 s) in the delay. Uninformative cues ("6," '7," "F," or "G") were presented as sensory control stimuli: presented early when the informative cue was late, and vice versa. Cue types (meaning the modality, timing, and how it was presented) were presented in a random order, with each type occurring an equal number of times. The delay from the end of stimulus presentation until the task onset lasted 26.375 s (6.775 s after stimulus presentation; 9.3 s after each cue). At the end of the delay (after 28.175 s in total), participants were presented with two test stimuli, which were a pure tone and a grating (0.8 s).

The task for participants was as follows: when the orientation sample was cued, they had to ignore the pitch stimulus and report whether this orientation was rotated clockwise or counterclockwise relative to the sample they had to remember. Likewise, when the pitch stimulus was cued, they had to report whether the tone was higher or lower than the sample they had to remember. Participants had 4 seconds to respond, and the intertrial interval was either 1.125 s or 3.575 s.

Experimental Design

Stimuli were centrally presented, which were used as both memory samples and targets. Visual stimuli were presented using a backprojection screen and auditory stimuli were presented using MRI compatible headphones. The presented sample was varied in orientation angle and pitch was varied randomly (±3.75° and ± 25 cents, approximate Gaussian) on a trial-by-trial basis. The distance of the target stimuli to the samples was also adjusted on a trial-by-trial basis, using a staircase procedure independently for the two tasks. This staircase was initialized during a training session outside of the scanner using shortened delays and slowly increasing task complexity to allow for more repetitions. For more

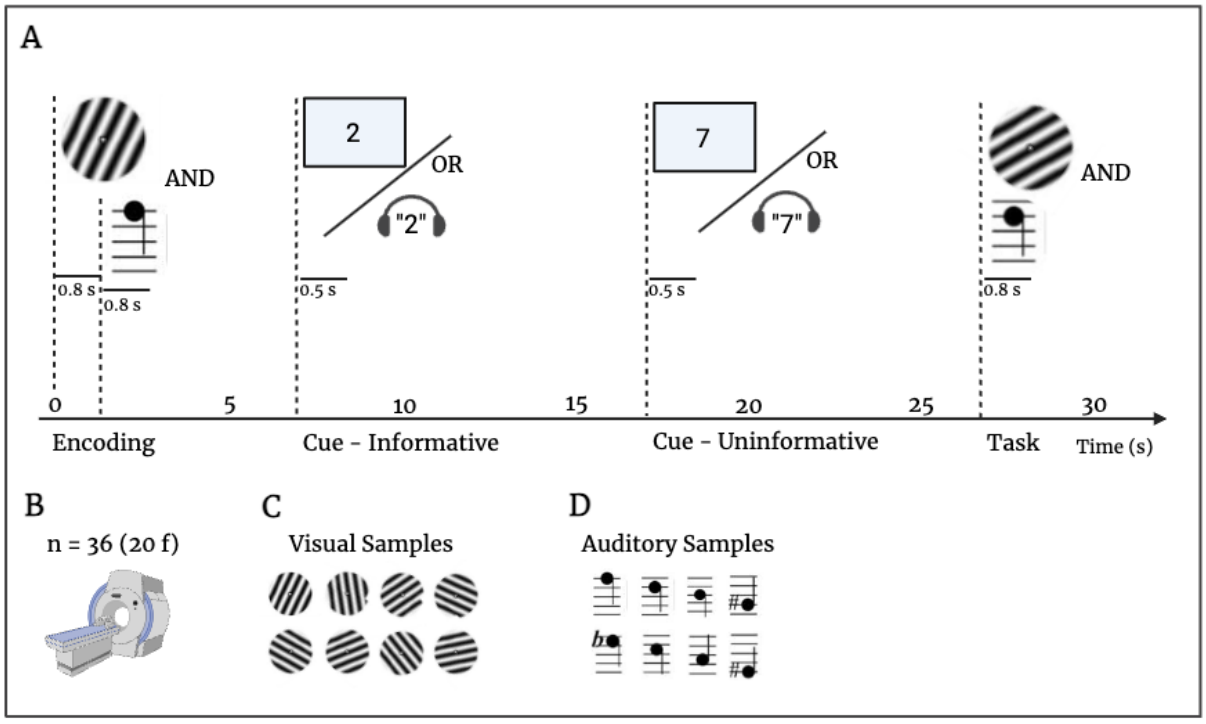

Figure 1. Experimental task. (A) **Schematic of the task design**. Participants memorized two stimuli, one visual grating and one auditory pure tone that were presented in random order. During the delay, participants were presented with an informative retro-cue indicating if the first or second of the two items had to be recalled at the end of the delay (making the other memory item irrelevant). This cue could be presented early or late in the delay. A second, uninformative cue served as a sensory control. Cues could be presented visually or auditorily. Unlike presented here, the uninformative cue could also be presented before the informative cue (in the case of an early uninformative cue with a late informative cue). At the end of the delay, another stimulus was presented visually or auditorily and participants were to decide whether this stimulus was rotated clockwise or counterclockwise (in the visual case) or whether the pitch was higher or lower (in the auditory case) in comparison to the stimulus they had encoded at the start of the trial. (B) **Participants and data acquisition**. (C) **Visual samples.** Eight different

rotations (sine-wave gratings: 187.5°, 217.5°, 323.5, 262.5°, 277.5°, 307.5°, 322.5°, and 352.5°) were used. (D) **Auditory samples.** Eight different pitches (pure tones: g♯" a♯", b", c♯", d", e", f", and g") were used.

details on the staircase procedure, see the original paper. Cues were either visually displayed centrally or auditorily presented using headphones. Participants took part in two sessions of four runs each. Each run consisted of 32 trials. The composition of these trials was a 2 (cued modality) × 2 (time of the informative cue) design. Within in this 2 × 2 design, there were eight trials. In each of these eight trials, the eight samples (eight orientations and eight tones) and all eight possible cues (cues indicating the first/second item, presented as numbers/letters, presented visually/auditorily) were used equally often. Cues and samples were randomly assigned to each other at the start of each run. Importantly, the cued modality or the remembered item had no predictive value over whether the informative cue indicated an item that was presented first or second. This means that a classifier with perfect content-specific information would have no selection-specific information, allowing us to carry out our analysis of interest without confounds.

Data collection

Functional MRI data were acquired on a 3 tesla MRI system (Siemens). In each run, 938 multiband EPI images were acquired (acceleration factor: 4; repetition time [TR]=1.225 s; 2×2×2 mm voxel size; 56 axial slices; 0.2-mm gap; 65° flip angle; 208-mm field of view). Presentation of the first sample of each trial was time-locked to the beginning of the acquisition of an image to minimize temporal variability. Additionally, a high-resolution, T1 weighted magnetization prepared rapid gradient echo anatomical volume was collected (192 sagittal slices; TR = 1.900 s; echo time = 2.52 ms; flip angle = 9°; field of view = 256 mm).

Preprocessing

Imaging data were processed using SPM12 (Friston et al., 1994) and CvCrossMANOVA (Allefeld & Haynes, 2014; Christophel et al., 2018). Functional data were motion corrected and anatomical data were registered to the functional images. For more details on preprocessing, see the original paper.

Multivariate Pattern Analyses

We used MVPA to investigate what neural mechanisms control whether an item stays in working memory or gets removed. MVPA provides a useful tool for identifying the neural regions that mediate working memory by focusing on changes in activation patterns as opposed to simply changes in the mean amplitude of the blood-oxygenation-level-dependent (BOLD) response. For this, we used CvCrossManova (Allefeld & Haynes, 2014), a technique for multivariate pattern analysis. Similar to traditional classifier-based MVPA analyses (for example Bettencourt & Xu, 2016; Christophel et al., 2012; Lee et al., 2013; Riggall & Postle, 2012; Serences et al., 2009), CvCrossManova can be used to identify representations held in neuronal activity patterns. Crucially, it has several advantages compared to traditional MVPA: CvCrossManova makes it possible to do a multi-class classification and is therefore not limited to binary classification, it performs a parameter-free analysis based on a probabilistic model of the data (the multivariate general linear model) and results in an interpretable multivariate effect size "*D*" (the pattern distinctness) (Allefeld & Haynes, 2014).

The effect size *D* tells how much of the variance can be explained by the effect encoded in our contrast. For the current study, that means we could define our contrast to differentiate between effects of an early and a late cue, whether the sample stimuli were presented visually or auditorily, and finally whether cues were presented visually or auditorily. Importantly, we could do this while separating the storage and control of working memory items. Note that the content of the stimuli, that is the orientation angles and pitches of the samples, as well as the

numbers and tones of the cues, were not deemed relevant in all current analyses. For all MVPAs, we used eightfold run-wise, (cross)-validation training on parameter estimates (beta-maps, in a run-based manner) for seven out of eight runs at the time and testing on the remaining run. This followed a leave-one-session-out (cross-)validation scheme. This was repeated until every run was used for testing.

To test the entirety of the recorded neural data for selection-specific activity in an unbiased fashion, we deployed a searchlight approach (Kriegeskorte et al., 2006). For a given voxel recorded from a given participant's brain, we defined a spherical cluster of voxels of in-brain voxel around this voxel (6 voxel radius). Then, we computed cross-classification accuracy across all cross-validation folds for the parcel of voxels within this cluster, and this accuracy value we stored in a 3D image at the position of the center of the cluster. This procedure was repeated until every brain voxel had been used as a cluster center. The resulting maps were smoothed and projected into MNI space using normalization parameters estimated during unified segmentation of the anatomical image. By doing this, the searchlight results could be statistically tested on the group-level. Statistical testing was performed using one-sided, one-sample t-tests in SPM. We used a voxel-level FWE correction to account for multiple comparisons (p(FWE)<.05, with a cluster-defining threshold of p<.001).

MVPA I: Keep versus Drop

To start off, our first MVPA focused on selection-specific activity that indirectly reflects the instruction conveyed by the cue: either remember in ("keep"), or removal from ("drop") working memory. The rationale behind this is that an early informative cue in the trial indirectly instructs participants to drop the sample that is not cued, whereas participants had to keep remembering both samples throughout the delay in case the cue was late. For example, when the first cue was an early informative cue (e.g., "1"), participants could drop the second

sample, since the late cue would be uninformative (e.g., "7"). Reversely, when the first cue was an uninformative cue (e.g., "6"), participants did not know yet which sample would be relevant for the upcoming task, so they still had to keep both items in memory throughout the delay, until the late informative cue told them which sample was relevant (e.g., "2").

We generated a general linear model with four regressors in each run: one regressor capturing the onset of the cue of auditory samples where the cue was early (representing "drop"), one for the cue of auditory samples where the cue was late (representing "keep"), one for the cue of visual samples where the cue was early ("drop"), and a fourth one for the cue of visual samples where the cue was late ("keep"). The onsets were convolved with a canonical hemodynamic response function. A fifth regressor capturing the mean of the run was added. Additionally, head motion parameters were added as regressors of no interest. The contrast for this analysis is a linear combination of the regression coefficients that are estimated from the analysis. The corresponding contrast for this model was as follows:

C_KeepDrop =

```
[  1        % early cue
   1        % early cue
  -1        % late cue
  -1  ]     % late cue
```

We used a classification approach by training every classifier on only one form of cues (e.g., early cues, meaning to drop one of the memory items) and then testing it on cues from trials where a different form was used (e.g., late cues, meaning to keep remembering both items). This way, we could see whether the control of keeping items in working memory generalizes to removing them, and if these activation patterns are dissociable. Because of the tight interplay between remembering and removing working memory items (Dames & Oberauer, 2022), we hypothesized that these patterns are not very dissociable. Note that in this analysis, we did not make use of *cross*-classification yet. Instead, we are merely checking whether we can decode

keep versus drop without differentiating between auditorily and visually presented cues or samples.

MVPA II: Control Analysis for Sample Modality

Our second MVPA focused on selection-specific activity that reflects the instruction of the cue, while controlling for the modality of the sample that is cued (visually or auditorily). The rationale behind visual versus auditory samples is that we wanted to rule out that the percept of the sample's modality might drive the effect. If that would be the case, the control of different modalities of sensory information, that is across auditory and visual information, might be represented differently in the brain.

We reused the model from our first MVPA but changed the contrast. This time, we were interested in comparing the difference in decoding between conditions where the sample modality was visual and conditions where the sample modality was auditory. Therefore, the contrast was as follows:

C_SampleModality

C_Visual =	C_Auditory =	
[0	[1	% early cue
1	0	% early cue
0	-1	% late cue
-1]	0]	% late cue

We used a cross-classification approach by training every classifier on only one form of cues (e.g., early cues for auditorily presented sample stimuli) and then testing it on cues from trials where a different form was used (e.g., late cues for visually presented sample stimuli), and vice-versa. From this, we wanted to decode keep versus drop while controlling for auditorily and visually presented samples. We hypothesized that by controlling for sample modality, we would see less pattern distinctness in the same areas compared to our first analysis. We

expected it would take away potential effects that might be caused by the percept of the modality of the sample stimuli.

MVPA III: Control Analysis for Sample and Cue Modality

For our final MVPA, we focused on selection-specific activity that reflects the instruction of the cue, this time controlling for the modality of the sample and the cue itself. The rationale behind visual versus auditory cues is that aside from sample modality, we also wanted to rule out that attentional processes towards visual and auditory information, as elicited by the cue presented visually or auditorily, might drive the effect. For our data, this means we are ignoring half of the trials: trials consisting of auditory samples with a visual cue, and trials of visual samples with auditory cues.

We generated a general linear model with eight regressors in each run: two regressors capturing the onset of early visual cues presented for visual and auditory samples, two for the onset of early auditory cues for visual and auditory samples, two for the onset of late visual cues for visual and auditory samples and two for the onset of late auditory cues for visual and auditory samples. The onsets were convolved with a canonical hemodynamic response function. A final regressor capturing the mean of the run was added. Additionally, head motion parameters were added as regressors of no interest. This time, we were interested in comparing the difference in decoding between conditions where the cue modality was visual and conditions where the cue modality was auditory. Therefore, the contrast was as follows:

C_BothModality

C_Visual =	C_Auditory =	
[0	[1	% aud sample early aud cue
0	0	% aud sample early vis cue
0	0	% vis sample early aud cue
1	0	% vis sample early vis cue
0	-1	% aud sample late aud cue
0	0	% aud sample late vis cue
0	0	% vis sample late aud cue
-1]	0]	% vis sample late vis cue

Although the current contrast could have been used for the first two analyses, we only decided to incorporate this analysis at a later stage. Therefore, we did not employ this contrast for the first two analyses at that time. We used a cross-classification approach by training every classifier on only one form of cues for a given meaning (e.g., auditory stimuli selected with an auditory cue), and then testing it on data from trials where a different form was used (e.g., visual stimuli selected with a visual cue), and vice-versa. This way, we aimed to decode drop versus keep while controlling for the percept of *both* the sample and the cue. We hypothesized that if the modality of the sample underlies some of the effect from the task, the modality of the cue would, too. Therefore, by controlling for sample and cue modality, we expected to see even less pattern distinctness in the same areas compared to our first two analyses.

Results

Multivariate pattern analyses of multiband fMRI data recorded throughout the 30-seconds-long trial were used to trace the remembering and removal of sample stimuli before and after informative cues. See the Methods section for details. Specifically, we wanted to see whether the remembering and removal of sample stimuli (as elicited by the retro-cue) rely upon a different neural substrate. We used multivariate encoding models tested with CvCrossManova (Allefeld & Haynes, 2014).

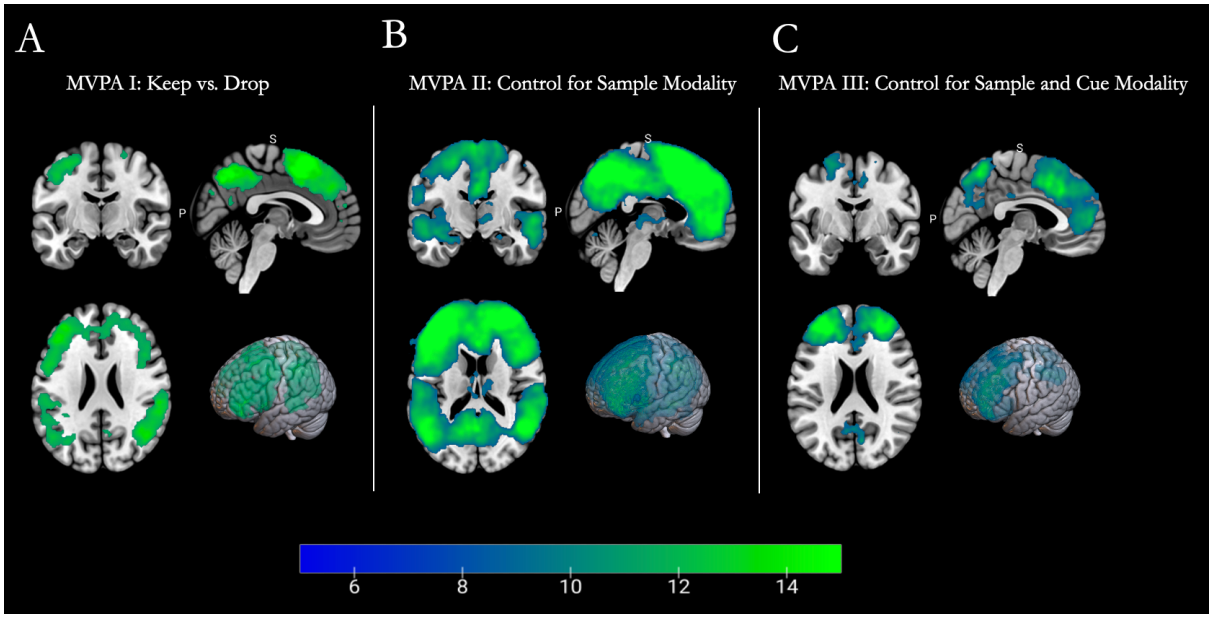

Figure 2. Results. (A) **MVPA I: Keep vs. Drop.** Regions Dissociable patterns of activity for keep vs. drop (forgetting vs. remembering) were observed mainly in the frontal and superior parietal cortex. (B) **MVPA II: Keep vs. Drop with Control for Sample Modality**. Dissociable patterns for keep vs. drop were observed mainly the frontal and parietal cortex, even after controlling for the modality of the sample. (C) **MVPA III: Keep vs. Drop with Control for Sample and Cue Modality**. Dissociable patterns for keep vs. drop were observed mainly the frontal and parietal cortex, even after controlling for the modality of the sample and of the cue. For all subfigures, results are rendered on a standard 3D brain template (display threshold: $p < .0000001$, uncorrected).

MVPA I and II: Keep versus Drop (+ Control Analysis for Sample Modality)

For our first two analyses, we wanted to see whether we could decode selection-specific activity that reflects the instruction conveyed by the cue (keep versus drop), while controlling for the modality of the sample in our second analysis. Retro-cues with different timings,

meaning whether the cue was early or late, were indirectly an instruction for participants to either drop or keep remembering the sample. For our second analysis, a control analysis, we distinguished between trials where the samples were presented visually and trials where samples were presented auditorily. A potential distinction between remembering and forgetting memory items could be consistent across different modalities of sensory samples. To avoid biasing our results based on prior assumptions about the localization of such selection-specific activity, we probed throughout the brain for cortical and noncortical regions encoding representations of the timing of the cue (MVPA I). For MVPA II we additionally distinguished between the different modalities of the sample. Using a searchlight-based decoding approach, we found that regions across all cortical lobe encoded whether the cue instructed the remembering or forgetting of memory representations for the rest of the trial: areas of at least the dorsolateral prefrontal cortex (dlPFC), ventrolateral prefrontal cortex (vlPFC), superior temporal gyrus (STG), intraparietal sulcus (IPS), and motor cortex (see Figure 2A, B). Specifically, regions of the medial frontal cortex (MFG) and posterior cingulate cortex (PCC) were the center of the largest cluster sizes in our first analysis (all p(FWE)<.05; MFG: MNI_{xyz} = [-46 50 4], t_{35} = 17.78, PCC: MNI_{xyz} = [34 -48 -28], t_{35} = 9.86, see Figure 2A). The same cortical areas showed activity even after controlling for sample modality (see Figure 2B), and activity seemed to be even more pronounced than in our first analysis. Specifically, regions of the superior frontal gyrus (SFG) were the center of the largest cluster sizes in this control analysis (all p(FWE) < .05; SFG: MNI_{xyz} = [-18 16 56], t_{71} = 22.25, see Figure 2B).

Results of these two analyses show dissociable patterns in the vast part of the cortical lobe, specifically the frontal and parietal cortex. These regions were previously discussed in the light of the dorsal ventroparietal attention network (DAN) (Szczepanski et al., 2013; Majerus et al., 2018; Rajan et al., 2021), the ventral frontoparietal attention network (VAN) (e.g., Corbetta & Shulman, 2002; Corbetta et al., 2008; Viviani, 2013) and the default mode

network (DMN) (e.g., Buckner et al., 2008; Leech & Smallwood, 2019; Pearson et al., 2011). Perhaps, selection-specific activity for remembering and forgetting items can coexist in multiple regions and depends on a whole network of activity rather than distinct areas. Alternatively, there might exist distinct areas that are involved in controlling the remembering or dropping items from working memory, and we simply do not find them. These areas might be difficult to pick up on with the current analysis, making them harder to pinpoint.

MVPA III: Control Analysis for Sample and Cue Modality

For our last control analysis, we wanted to see whether we could decode selection-specific activity for keep versus drop while considering the modality of *both* the sample and the cue. This was done as a control for attentional processes towards the sample and cue. We again probed throughout the brain for cortical and noncortical regions. This time we decoded keep versus drop while considering the representations of the modality of the sample and of the cue. Thereby we disregarded half of our trials (e.g., trials where the sample was auditory and the cue was visual, and vice-versa). Using a searchlight-based decoding approach, we found again that areas across all cortical lobes encoded whether the cue instructed the remembering or forgetting of memory representations for the rest of the trial (see Figure 2C). Regions of the medial frontal gyrus (MFG) and posterior cingulate cortex (PCC) were the center of the largest cluster sizes in this analysis (all p(FWE) <.05; MFG: MNI_{xyz} = [-26 44 26], t_{71} = 19.09, PCC: MNI_{xyz} = [-42 -34 -46], t_{71} = 15.04, see Figure 2C).

This time, although still a lot of regions across all cortical lobes seem to encode whether the cue instructed the remembering or forgetting of memory representations for the rest of the trial, controlling for sample and cue modality seems to have had some effect and lowered the decoding considerably. Specifically, decoding in the STG, an area known for its involvement in auditory processing (e.g., Bigler et al., 2007; Mesgarani et al., 2014; Wernicke, 1874), seems

to have reduced. Nonetheless, it remains the case that even after we control for sample and cue modality, the amount of information that we see in the brain in relation to the current task is substantial. Similar to our first two analyses, we found many regions of the frontal and parietal cortex to show dissociable patterns in response to the modality of the cue.

Discussion

In the current study, we used multivariate pattern analyses to investigate what neural mechanisms control whether an item stays in working memory or gets removed. Across all analyses, vast parts of the cortical lobe, specifically the frontal and parietal cortex show dissociable patterns in response to the delay period in the retro-cue task. Multiple attention networks such as the dorsal frontoparietal network, the ventral frontoparietal network, and the default mode network show encoding of the remembering and forgetting of working memory representations, albeit with slightly different amplitudes. These results suggest that multiple networks might underlie the control of the remembering and removal of items in working memory simultaneously.

Networks

The default attention network has been implicated in attentional control of different types of attention, including spatial, feature, and object attention (Morishima et al., 2009; Slagter et al., 2007). This top-down control of attention arises chiefly in the frontal and parietal cortex (Szczepanski et al., 2013; Majerus et al., 2018; Rajan et al., 2021), of which areas FEF, SEF, IPS1-5 and SPL1 specifically have shown to be the main areas of the DAN (Szczepanski et al., 2013). The DAN is responsible for the selection of the appropriate response or action that is necessary for the orientation of attention (Majerus et al., 2016; Majerus et al., 2018; Szczepanski et al., 2013), and it has even been shown that it encodes information about trial-specific instructions (Riggall & Postle, 2012). This indicates that the DAN may exert a more general role of task-related attentional control, and it is likely that we are seeing activity related to task instructions in the current retro-cue task.

The ventral attention network responds along with the DAN when behaviorally relevant targets are detected (e.g., Corbetta et al., 2008; Corbetta et al., 2000). The interaction of the

DAN and VAN has been proposed to be as follows: when attention is reoriented to a new source of information, output from the VAN interrupts the ongoing selection in the DAN, which in turn shifts attention toward the novel object of interest (Corbetta and Shulman, 2002). In the current task, possibly it is the case that as soon as the informative cue is presented, the VAN is involved in shifting attention toward the behaviorally relevant cue, so that it can be used to complete the upcoming task. Thus, the patterns observed in the current retro-cue task may be driven by the interaction between *both* attention networks.

The default mode network, a set of distributed regions in the parietal, temporal and frontal cortex, shows reductions in activity during attention-demanding tasks (Smallwood et al., 2021). However, an active role for the DMN has been suggested as well. It has been suggested that the default mode network could be associated with retrieval of information, especially from long-term memory, as well (Kim et al., 2010; Leech & Sharp, 2014; Pearson et al., 2011). Therefore, involvement of the DMN during the delay in the current retro-cue task might represent the active retrieval of the sample's memory representations.

It is possible that the control of the remembering and removal from items in working memory is such a cognitively demanding and therefore complicated operation that it requires activation of multiple networks. Therefore, it might extend to many areas of the human brain. However, we should note that a vast part of the cortical lobe showed dissociable patterns, and not exclusively in the previously mentioned attention networks. While we identified some networks that could underlie the activation observed in our task, the extensive amount of brain activation means these results should be interpreted with caution. However, the Salience Network - involved in interoception of feelings associated with reward - is not obviously visible in the current results. That means that at the least, it is not the case that *all* possible networks are associated with our task. Altogether, although our goal was to study memory control at the simplest level and pinpoint specific brain areas that might be involved, selection-specific

activity might simply not be restricted to very distinct areas of the brain. Future research is needed to clarify these widespread activations.

Limitations and Future Directions

It is possible that specific brain areas that harbor representations of selection-specific activity are invisible to our research method or analyses. Memory control processes might use a neural code that is independent of external properties such as the modality of the sample or cue. In the dorsal attention network, there exist functional microstructures that are specific for controlling attention based on the specific information to be attended (Rajan et al., 2021). This suggests that indeed our research method or analyses might simply not be discriminative enough to find dissociable patterns related to specific structures. Therefore, future work could deploy regions of interest (ROIs) to target specific areas in the brain. ROIs could be drawn around the FEF, SEF, IPS1-5 and SPL1, which are well-known DAN areas. This way, a more detailed picture of the current findings can be painted.

Perhaps, our two control analyses (MVPA II & III) did not target what we expected. Our expectation was that controlling for sample modality would take away the effect that would be driven by the percept of the modality of the sample. This would then lower the overall selection-specific information in the results. Unlike our expectation, the first control analysis (MVPA II) showed *more* pronounced selection-specific information after controlling for the sample modality, rather than less. While the analysis was conducted with attention to detail, it is always possible that there is an error or discrepancy within the pipeline. Therefore, future research may benefit from additional validation measures or alternative approaches to ensure the reliability and robustness of the current findings. Specifically, it may be interesting to opt for a different but similar multivariate pattern analysis method, such as the Decoding Toolbox (Hebart et al., 2015) to see how the results compare. At this stage we are unsure why the first

control analysis showed more selection-specific information compared to our first analysis (MVPA I). For this reason we refrain from drawing conclusions about its results.

Concerning the interpretations of the results altogether, there are a few more caveats to consider. First of all, the causal chain of events cannot be inferred from these analyses. fMRI analyses present correlational results, not causal. Therefore, it is impossible to distinguish if the patterns we see are the *cause* of forgetting or remembering working memory items, or the *effect*. Relatedly, it is unclear whether the different networks we see in our data encode the control of working memory independently of one another, or if it is one massive network working together instead. All in all, we should not infer any causal relationships from the current results. A different caveat is that it is difficult to determine which working memory process is driving the effect. We do not know for certain if the effects truly come from the remembering and forgetting of working memory items. When instructed to forget one of the working memory items, it is possible that continuing to remember the other item is driving the effect instead. Perhaps, remembering versus some baseline would give similar results. In summary, while our fMRI analyses reveal dissociable patterns of working memory processes, it is impossible to determine what the causal chain of events looks like, how the interaction between the various cortical areas works, and exactly which memory processes are driving these effects.

Conclusion

While the current results are different from what we initially hypothesized, they nevertheless present interesting findings. The control of the remembering in and removal of items from working memory appears to be a more complicated and intricate process that extends beyond distinct brain areas or networks. Even then, this process might not be restricted to the networks mentioned in the current paper, indicating a broader neural involvement in the brain's decision-

making process. At least the dorsal frontoparietal, ventral frontoparietal, and default mode network seem to be involved when we perform a new memory action (in our case: “drop” from working memory) or continue to perform the same action (“keep” in working memory). Even after controlling for the percept of the modality of the sample and cue, the amount of dissociable patterns we see in the brain is massive. However, the current analyses have drawbacks, and we should interpret these results with caution. Our findings suggest that even the seemingly smallest operation is extremely complex in the brain. Perhaps, the executive control in the brain is distributed across multiple regions, much like the complex and distributed nature of storing working memory contents. This highlights the need for further research to understand the neural mechanisms underlying working memory control.

References

Allefeld, C., & Haynes, J. D. (2014). Searchlight-based multi-voxel pattern analysis of fMRI by cross-validated MANOVA. *Neuroimage, 89*, 345-357. https://doi.org/10.1016/j.neuroimage.2013.11.043

Antono, J. E., Dang, S., Auksztulewicz, R., & Pooresmaeili, A. (2023). Distinct patterns of connectivity between brain regions underlie the intra-modal and cross-modal value-driven modulations of the visual cortex. *Journal of Neuroscience, 43*(44), 7361-7375. DOI: 10.1523/JNEUROSCI.0355-23.2023

Arbib, M. A. (Ed.). (2003). *The handbook of brain theory and neural networks*. MIT press.

Awh, E. & Jonides, J. (2001) Overlapping mechanisms of attention and spatial working memory. *Trends Cogn. Sci. 5*, 119–126. https://doi.org/10.1016/S1364-6613(00)01593-X

Baddeley, A. (1992). Working memory. *Science*, *255*, 556-559. https://doi.org/10.1126/science.1736359

Bettencourt, K. C., & Xu, Y. (2016). Decoding the content of visual short-term memory under distraction in occipital and parietal areas. *Nature Neuroscience, 19*, 150–157. https://doi.org/10.1038/nn.4174

Bigler, E. D., Mortensen, S., Neeley, E. S., Ozonoff, S., Krasny, L., Johnson, M., ... & Lainhart, J. E. (2007). Superior temporal gyrus, language function, and autism. *Developmental neuropsychology, 31*(2), 217-238. DOI: 10.1080/87565640701190841

Bisley, J. W., Zaksas, D., Droll, J. A., & Pasternak, T. (2004). Activity of neurons in cortical area MT during a memory for motion task. *Journal of neurophysiology, 91*(1), 286-300. DOI: 10.1152/jn.00870.2003.

Björk, R. A. (1970). Positive forgetting: The noninterference of items intentionally forgotten. *Journal of Verbal Learning and Verbal Behavior, 9*(3), 255-268. https://doi.org/10.1016/S0022-5371(70)80059-7

Buckner, R. L., Andrews-Hanna, J. R., & Schacter, D. L. (2008). The brain's default network: anatomy, function, and relevance to disease. *Annals of the new York Academy of Sciences, 1124*(1), 1-38. DOI: 10.1196/annals.1440.011

Chelazzi, L., Miller, E. K., Duncan, J., & Desimone, R. (2001). Responses of neurons in macaque area V4 during memory-guided visual search. *Cerebral cortex, 11*(8), 761-772. DOI: 10.1093/cercor/11.8.761

Christophel, T., Endisch, C., & Haynes, J. D. (2013). Decoding invariant representations in visual working memory. *Journal of Vision, 13*(9), 927-927. https://doi.org/10.1167/13.9.927

Christophel, T. B., Iamshchinina, P., Yan, C., Allefeld, C., & Haynes, J.-D. (2018). Cortical specialization for attended versus unattended working memory. *Nature Neuroscience, 21*, 494–496. https://doi.org/10.1038/s41593-018-0094-4

Christophel, T. B., Klink, P. C., Spitzer, B., Roelfsema, P. R., & Haynes, J. D. (2017). The distributed nature of working memory. *Trends in cognitive sciences, 21*(2), 111-124. https://doi.org/10.1016/j.tics.2016.12.007

Christophel, T. B., Hebart, M. N., & Haynes, J.D. (2012). Decoding the contents of visual short-term memory from human visual and parietal cortex. *Journal of Neuroscience, 32*, 12983–12989. https://doi.org/10.1523/JNEUROSCI.0184-12.2012

Christophel, T., Weber, S., Yan, C., Stopak, L., Hetzer, S., & Haynes, J. D. (2024). Nonfrontal Control of Working Memory. *Journal of Cognitive Neuroscience, 1*-11. https://doi.org/10.1162/jocn_a_02127

Collette, F., & Van der Linden, M. (2002). Brain imaging of the central executive component of working memory. *Neuroscience & Biobehavioral Reviews, 26*(2), 105-125. https://doi.org/10.1016/S0149-7634(01)00063-X

Constantinidis, C., & Klingberg, T. (2016). The neuroscience of working memory capacity and training. *Nature Reviews. Neuroscience, 17*, 438–449. https://doi.org/10.1038/nrn.2016.43

Corbetta, M., Patel, G., & Shulman, G. L. (2008). The reorienting system of the human brain: from environment to theory of mind. *Neuron, 58*(3), 306-324. https://doi.org/10.1016/j.neuron.2008.04.017.

Corbetta, M., & Shulman, G. L. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nature reviews neuroscience, 3*(3), 201-215. DOI: 10.1038/nrn755

Cowan, N. (1998). Attention and memory: An integrated framework. Oxford University Press.

Cowan, N. (1988). Evolving conceptions of memory storage, selective attention, and their mutual constraints within the human information-processing system. *Psychological bulletin, 104*(2), 163.

Cowan, N. (2010). The magical mystery four: How is working memory capacity limited, and why?. *Current directions in psychological science, 19*(1), 51-57. https://doi.org/10.1177/0963721409359277

Cowan, N., Elliott, E. M., Saults, J. S., Morey, C. C., Mattox, S., Hismjatullina, A., & Conway, A. R. (2005). On the capacity of attention: Its estimation and its role in working memory and cognitive aptitudes. *Cognitive psychology, 51*(1), 42-100. https://doi.org/10.1016/j.cogpsych.2004.12.001

Curtis, C. E., & D'Esposito, M. (2003). Persistent activity in the prefrontal cortex during working memory. *Trends in cognitive sciences, 7*(9), 415-423. https://doi.org/10.1016/S1364-6613(03)00197-9

Dames, H., & Oberauer, K. (2022). Directed forgetting in working memory. Journal of *Experimental Psychology: General, 151*(12), 2990. DOI: 10.1037/xge0001256

Davey, J., Thompson, H. E., Hallam, G., Karapanagiotidis, T., Murphy, C., De Caso, I., ... & Jefferies, E. (2016). Exploring the role of the posterior middle temporal gyrus in semantic cognition: Integration of anterior temporal lobe with executive processes. *Neuroimage, 137*, 165-177. https://doi.org/10.1016/j.neuroimage.2016.05.051

d'Esposito, M. (2007). From cognitive to neural models of working memory. *Philosophical Transactions of the Royal Society B: Biological Sciences, 362*(1481), 761-772. https://doi.org/10.1098/rstb.2007.2086

Ecker, U. K., Lewandowsky, S., & Oberauer, K. (2014). Removal of information from working memory: A specific updating process. *Journal of Memory and Language, 74*, 77-90. https://doi.org/10.1016/j.jml.2013.09.003

Emrich, S.M. (2015) What are the roles of sensory and parietal activity in visual short-term memory? *Mechanisms of Sensory Working Memory: Attention and Perfomance XXV. Elsevier*. http://dx.doi.org/10.1016/B978-0-12-801371-7.00006-5

Ester, E. F., Sprague, T. C., & Serences, J. T. (2015). Parietal and frontal cortex encode stimulus-specific mnemonic representations during visual working memory. *Neuron*, *87*(4), 893-905. https://doi.org/10.1016/j.neuron.2015.07.013

Friston, K. J., Holmes, A. P., Worsley, K. J., Poline, J. P., Frith, C. D., & Frackowiak, R. S. (1994). Statistical parametric maps in functional imaging: a general linear approach. *Human brain mapping, 2*(4), 189-210. DOI: 10.1002/hbm.460020402

Funahashi, S. (2017). Working memory in the prefrontal cortex. *Brain sciences, 7*(5), 49. https://doi.org/10.3390/brainsci7050049

Fuster, J.M. (1995) Memory in the Cerebral Cortex: An Empirical Approach to Neural Networks in the Human and Nonhuman Primate, *MIT Press*

Fuster, J.M. and Jervey, J.P. (1982) Neuronal firing in the inferotemporal cortex of the monkey in a visual memory task. *J. Neurosci. 2*, 361–375. DOI: 10.1523/JNEUROSCI.02-03-00361.1982

Fuster, J. and Jervey, J. (1981) Inferotemporal neurons distinguish and retain behaviorally relevant features of visual stimuli. *Science 212*, 952–955. DOI: 10.1126/science.7233192

Gillebert, C. R., Mantini, D., Thijs, V., Sunaert, S., Dupont, P., & Vandenberghe, R. (2011). Lesion evidence for the critical role of the intraparietal sulcus in spatial attention. *Brain, 134*(6), 1694-1709. DOI: 10.1093/brain/awr085

Goldman-Rakic, P. (1995) Cellular basis of working memory. *Neuron, 14*, 477–485 https://doi.org/10.1016/0896-6273(95)90304-6

Gunseli, E., van Moorselaar, D., Meeter, M., & Olivers, C. N. (2015). The reliability of retro-cues determines the fate of noncued visual working memory representations. *Psychonomic bulletin & review, 22*, 1334-1341. DOI: 10.3758/s13423-014-0796-x

Harrison, S.A. and Tong, F. (2009) Decoding reveals the contents of visual working memory in early visual areas. *Nature, 458*, 632–635. https://doi.org/10.1038/nature07832

Hasher, L., Zacks, R. T., & May, C. P. (1999). Inhibitory control, circadian arousal, and age. DOI: 10.7551/mitpress/1480.001.0001

Hebart, M. N., Görgen, K., & Haynes, J. D. (2015). The Decoding Toolbox (TDT): a versatile software package for multivariate analyses of functional imaging data. *Frontiers in neuroinformatics, 8*, 88. https://doi.org/10.3389/fninf.2014.00088

Jacobsen, C.F. (1936) The functions of the frontal association areas in monkeys. *Comparative Psychology Monographs 13,* 1–60

Jonides, J., & Nee, D. E. (2006). Brain mechanisms of proactive interference in working memory. *Neuroscience, 139*(1), 181-193. https://doi.org/10.1016/j.neuroscience.2005.06.042

Kriegeskorte, N., Goebel, R., & Bandettini, P. (2006). Information-based functional brain mapping. *Proceedings of the National Academy of Sciences, U.S.A., 103*, 3863–3868. https://doi.org/10.1073/pnas.0600244103

Kim, H., Daselaar, S. M., & Cabeza, R. (2010). Overlapping brain activity between episodic memory encoding and retrieval: roles of the task-positive and task-negative networks. *Neuroimage, 49*(1), 1045-1054. https://doi.org/10.1016/j.neuroimage.2009.07.058

Kumar, S., Joseph, S., Gander, P. E., Barascud, N., Halpern, A. R., & Griffiths, T. D. (2016). A brain system for auditory working memory. *Journal of Neuroscience, 36(*16), 4492-4505. DOI: 10.1523/JNEUROSCI.4341-14.2016

Lara, A.H. and Wallis, J.D. (2014) Executive control processes underlying multi-item working memory. *Nat. Neurosci. 17*, 876–883. DOI: 10.1038/nn.3702

Lee, S.-H., Kravitz, D. J., & Baker, C. I. (2013). Goal-dependent dissociation of visual and prefrontal cortices during working memory. *Nature Neuroscience, 16*, 997–999. https://doi.org/10.1038/nn.3452

Leech, R., & Sharp, D. J. (2014). The role of the posterior cingulate cortex in cognition and disease. *Brain, 137*(1), 12-32. DOI: 10.1093/brain/awt162

Leech, R., & Smallwood, J. (2019). The posterior cingulate cortex: Insights from structure and function. *Handbook of clinical neurology, 166*, 73-85. https://doi.org/10.1016/B978-0-444-64196-0.00005-4

Linke, A. C., Vicente-Grabovetsky, A., Mitchell, D. J., & Cusack, R. (2011). Encoding strategy accounts for individual differences in change detection measures of VSTM. *Neuropsychologia, 49*(6), 1476-1486. https://doi.org/10.1016/j.neuropsychologia.2010.11.034

Ma, W. J., Husain, M., & Bays, P. M. (2014). Changing concepts of working memory. *Nature Neuroscience, 17*, 347–356. https://doi.org/10.1038/nn.3655

Majerus, S., Cowan, N., Péters, F., Van Calster, L., Phillips, C., & Schrouff, J. (2016). Cross-modal decoding of neural patterns associated with working memory: Evidence for attention-based accounts of working memory. *Cerebral Cortex, 26*(1), 166-179. DOI: 10.1093/cercor/bhu189

Majerus, S., Péters, F., Bouffier, M., Cowan, N., & Phillips, C. (2018). The dorsal attention network reflects both encoding load and top–down control during working memory. *Journal of Cognitive Neuroscience, 30*(2), 144-159. DOI: 10.1162/jocn_a_01195

Marshuetz, C., Smith, E. E., Jonides, J., DeGutis, J., & Chenevert, T. L. (2000). Order information in working memory: fMRI evidence for parietal and prefrontal mechanisms. *Journal of cognitive neuroscience, 12*(2), 130-144. https://doi.org/10.1162/08989290051137459

Mendoza-Halliday, D., Torres, S., & Martinez-Trujillo, J. C. (2014). Sharp emergence of feature-selective sustained activity along the dorsal visual pathway. *Nature neuroscience, 17*(9), 1255-1262. https://doi.org/10.1038/nn.3785

Mendoza-Halliday, D., Torres, S., & Martinez-Trujillo, J. (2015). Working memory representations of visual motion along the primate dorsal visual pathway. *Mechanisms of Sensory Working Memory: Attention and Perfomance XXV*. http://dx.doi.org/10.1016/B978-0-12-811042-3.00012-8

Mesgarani, N., Cheung, C., Johnson, K., & Chang, E. F. (2014). Phonetic feature encoding in human superior temporal gyrus. *Science, 343*(6174), 1006-1010. DOI: 10.1126/science.1245994

Mesulam, M. M. (1998). From sensation to cognition. *Brain: a journal of neurology, 121*(6), 1013-1052. DOI: 10.1093/brain/121.6.1013

Miller, E. K., Li, L., & Desimone, R. (1993). Activity of neurons in anterior inferior temporal cortex during a short-term memory task. *Journal of Neuroscience, 13*(4), 1460-1478. DOI: 10.1523/JNEUROSCI.13-04-01460.1993

Miyashita, Y. & Chang, H.S. (1988) Neuronal correlate of pictorial short-term memory in the primate temporal cortex. *Nature, 331*, 68–70. DOI: 10.1038/331068a0

Morishima, Y., Akaishi, R., Yamada, Y., Okuda, J., Toma, K., & Sakai, K. (2009). Task-specific signal transmission from prefrontal cortex in visual selective attention. *Nature Neuroscience, 12*, 85–91. DOI: https://doi.org/10.1038/nn.2237

Narayanan, N. S., Prabhakaran, V., Bunge, S. A., Christoff, K., Fine, E. M., & Gabrieli, J. D. (2005). The role of the prefrontal cortex in the maintenance of verbal working memory: an event-related FMRI analysis. *Neuropsychology, 19*(2), 223. https://doi.org/10.1037/0894-4105.19.2.223

Nee, D. E., & Jonides, J. (2009). Common and distinct neural correlates of perceptual and memorial selection. *Neuroimage, 45*(3), 963-975. DOI: 10.1016/j.neuroimage.2009.01.005

Nobre, A. C., Coull, J. T., Maquet, P., Frith, C. D., Vandenberghe, R., & Mesulam, M. M. (2004). Orienting attention to locations in perceptual versus mental representations. *Journal of cognitive neuroscience, 16*(3), 363-373. DOI: 10.1162/089892904322926700

Oberauer, K. (2009). Design for a working memory. *Psychology of learning and motivation, 51*, 45-100. https://doi.org/10.1016/S0079-7421(09)51002-X

Oberauer, K. (2001). Removing irrelevant information from working memory: a cognitive aging study with the modified Sternberg task. *Journal of experimental Psychology: learning, Memory, and cognition, 27*(4), 948. https://doi.org/10.1037/0278-7393.27.4.948

Oberauer, K. (2018). Removal of irrelevant information from working memory: Sometimes fast, sometimes slow, and sometimes not at all. *Annals of the New York Academy of Sciences, 1424(*1), 239-255. DOI: 10.1111/nyas.13603

Oberauer, K., Lewandowsky, S., Farrell, S., Jarrold, C., & Greaves, M. (2012). Modeling working memory: An interference model of complex span. *Psychonomic Bulletin & Review, 19,* 779–819. https://doi.org/10.3758/s13423-012-0272-4

Panichello, M. F., & Buschman, T. J. (2021). Shared mechanisms underlie the control of working memory and attention. *Nature, 592*, 601–605. https://doi.org/10.1038/s41586-021-03390-w

Pasternak, T., Lui, L. L., & Spinelli, P. M. (2015). Unilateral prefrontal lesions impair memory-guided comparisons of contralateral visual motion. *Journal of Neuroscience, 35*(18), 7095-7105. DOI: 10.1523/JNEUROSCI.5265-14.2015

Pearson, J. M., Heilbronner, S. R., Barack, D. L., Hayden, B. Y., & Platt, M. L. (2011). Posterior cingulate cortex: adapting behavior to a changing world. *Trends in cognitive sciences, 15*(4), 143-151. https://doi.org/10.1016/j.tics.2011.02.002

Petersson, K. M., Gisselgård, J., Gretzer, M., & Ingvar, M. (2006). Interaction between a verbal working memory network and the medial temporal lobe. *Neuroimage, 33*(4), 1207-1217. https://doi.org/10.1016/j.neuroimage.2006.07.042

Postle, B. R. (2006). Working memory as an emergent property of the mind and brain. Neuroscience, 139(1), 23-38. https://doi.org/10.1016/j.neuroscience.2005.06.005

Pratte, M. S., & Tong, F. (2014). Spatial specificity of working memory representations in the early visual cortex. Journal of vision, 14(3), 22-22. https://doi.org/10.1167/14.3.22

Quentin, R., King, J. R., Sallard, E., Fishman, N., Thompson, R., Buch, E. R., & Cohen, L. G. (2019). Differential brain mechanisms of selection and maintenance of information during working memory. *Journal of Neuroscience, 39*(19), 3728-3740. DOI: 10.1523/JNEUROSCI.2764-18.2019

Rajan, A., Meyyappan, S., Liu, Y., Samuel, I. B. H., Nandi, B., Mangun, G. R., & Ding, M. (2021). The microstructure of attentional control in the dorsal attention network. *Journal of cognitive neuroscience, 33*(6), 965-983. DOI: 10.1162/jocn_a_01710

Riggall, A. C., & Postle, B. R. (2012). The relationship between working memory storage and elevated activity as measured with functional magnetic resonance imaging. *Journal of Neuroscience, 32*, 12990–12998. https://doi.org/10.1523/JNEUROSCI.1892-12.2012

Riley, M.R. & Constantinidis, C. (2016) Role of prefrontal persistent activity in working memory. *Front. Syst. Neurosci. 9,* 181. https://doi.org/10.3389/fnsys.2015.00181

Szczepanski, S. M., Pinsk, M. A., Douglas, M. M., Kastner, S., & Saalmann, Y. B. (2013). Functional and structural architecture of the human dorsal frontoparietal attention network. Proceedings of the National Academy of Sciences, 110(39), 15806-15811. DOI: 10.1073/pnas.1313903110

Serences, J. T. (2016). Neural mechanisms of information storage in visual short-term memory. *Vision research*, *128*, 53-67. https://doi.org/10.1016/j.visres.2016.09.010

Serences, J. T., Ester, E. F., Vogel, E. K., & Awh, E. (2009). Stimulus-specific delay activity in human primary visual cortex. *Psychological Science, 20*, 207–214. https://doi.org/10.1111/j.1467-9280.2009.02276.x

Slagter, H. A., Giesbrecht, B., Kok, A., Weissman, D. H., Kenemans, J. L., Woldorff, M. G., & Mangun, G. R. (2007). fMRI evidence for both generalized and specialized components of attentional control. *Brain research, 1177*, 90-102. https://doi.org/10.1016/j.brainres.2007.07.097

Smallwood, J., Bernhardt, B. C., Leech, R., Bzdok, D., Jefferies, E., & Margulies, D. S. (2021). The default mode network in cognition: a topographical perspective. *Nature reviews neuroscience, 22*(8), 503-513. DOI: 10.1038/s41583-021-00474-4

Smith, E.E., & Jonides, J. (1997). Working memory: A view from neuroimaging. *Cognitive Psychology, 33*(1), 5-42. https://doi.org/10.1006/cogp.1997.0658

Souza, A.S. & K. Oberauer. 2016. In search of the focus of attention in working memory: 13 years of the retro-cue effect. *Atten. Percept. Psychophys. 78*: 1839–1860. DOI: 10.3758/s13414-016-1108-5

Sreenivasan, K. K., Curtis, C. E., & D'Esposito, M. (2014). Revisiting the role of persistent neural activity during working memory. *Trends in cognitive sciences, 18*(2), 82-89. https://doi.org/10.1016/j.tics.2013.12.001

Supèr, H., Spekreijse, H., & Lamme, V. A. (2001). A neural correlate of working memory in the monkey primary visual cortex. *Science, 293*(5527), 120-124. DOI: 10.1126/science.1060496

Szczepanski, S. M., Pinsk, M. A., Douglas, M. M., Kastner, S., & Saalmann, Y. B. (2013). Functional and structural architecture of the human dorsal frontoparietal attention network. *Proceedings of the National Academy of Sciences, 110*(39), 15806-15811. DOI: 10.1073/pnas.1313903110

van Moorselaar, D., Olivers, C. N., Theeuwes, J., Lamme, V. A., & Sligte, I. G. (2015). Forgotten but not gone: Retro-cue costs and benefits in a double-cueing paradigm suggest multiple states in visual short-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 41*(6), 1755. DOI: 10.1037/xlm0000124

Viviani, R. (2013). Emotion regulation, attention to emotion, and the ventral attentional network. *Frontiers in human neuroscience, 7*, 746. https://doi.org/10.3389/fnhum.2013.00746

Wernicke, C. (1874*). Der aphasische Symptomencomplex: eine psychologische Studie auf anatomischer Basis*. Cohn & Weigert.

Wongupparaj, P., Kumari, V., & Morris, R. G. (2015). The relation between a multicomponent working memory and intelligence: The roles of central executive and short-term storage functions. *Intelligence, 53*, 166-180. https://doi.org/10.1016/j.intell.2015.10.007

Wu, Y. H., Uluç, I., Schmidt, T. T., Tertel, K., Kirilina, E., & Blankenburg, F. (2018). Overlapping frontoparietal networks for tactile and visual parametric working memory representations. *Neuroimage, 166*, 325-334. https://doi.org/10.1016/j.neuroimage.2017.10.059

Xu, Y. & Jeong, S.K. (2015) The contribution of human superior intraparietal sulcus to visual short-term memory and perception. *Mechanisms of Sensory Working Memory: Attention and Perfomance XXV*. Elsevier. http://dx.doi.org/10.1016/B978-0-12-801371-7.00004-1

Zhou, Y., Curtis, C. E., Sreenivasan, K. K., & Fougnie, D. (2022). Common neural mechanisms control attention and working memory. *Journal of Neuroscience, 42*, 7110–7120. https://doi.org/10.1523/JNEUROSCI.0443-22.2022