Toward a model-based cognitive neuroscience of working memory subprocesses

Boag, R.J.; Miletić, S.; Trutti, A.C.; Forstmann, B.U.; Turner, B.M.

Toward a Model-Based Cognitive Neuroscience of Working Memory Subprocesses

Russell J. Boag, Steven Miletić, Anne C. Trutti, and Birte U. Forstmann

Abstract Working memory (WM) refers to a set of processes that makes task-relevant information accessible to higher-level cognitive processes including abstract reasoning, decision-making, learning, and reading comprehension. In this chapter, we introduce the concept of WM and outline key behavioral and neural evidence for a number of critical subprocesses that support WM and which have become recent targets of cognitive neuroscience. We discuss common approaches to linking brain and behavior in WM research seeking to identify the neural basis of WM subprocesses. We draw attention to limitations of common approaches and suggest that much progress could be made by applying several of the recent methodological advances in model-based cognitive neuroscience discussed throughout this book (see Chapters "An Introduction to EEG/MEG for Model-Based Cognitive Neuroscience", "Ultra-High Field Magnetic Resonance Imaging for Model-Based Neuroscience", "Advancements in Joint Modeling of Neural and Behavioral Data", "Cognitive Models as a Tool to Link Decision Behavior With EEG Signals", and "Linking Models with Brain Measures"). Overall, the purpose of this chapter is to give a broad overview of WM as seen through the lens of model-based cognitive neuroscience and to summarize our current state of knowledge of WM subprocesses and their neural basis. We hope to outline a path forward to a more complete neurocomputational understanding of WM.

Keywords Working memory · Neural basis of memory · Subcortical circuits

R. J. Boag (✉) · S. Miletić · B. U. Forstmann
Integrative Model-Based Cognitive Neuroscience Research Unit, University of Amsterdam, Amsterdam, The Netherlands
e-mail: r.j.boag@uva.nl

A. C. Trutti
Integrative Model-Based Cognitive Neuroscience Research Unit, University of Amsterdam, Amsterdam, The Netherlands

Cognitive Psychology Unit, Institute of Psychology & Leiden Institute for Brain and Cognition, Leiden University, Leiden, The Netherlands

B. U. Forstmann, B. M. Turner (eds.), *An Introduction to Model-Based Cognitive Neuroscience*, https://doi.org/10.1007/978-3-031-45271-0_11

1 Introduction

1.1 Working Memory

The acronym "WM" stands for *working memory*. You have just encoded a representation – "WM" – of an abstract concept – "working memory" – into a distinct neural assembly in your own WM, probably somewhere in your prefrontal cortex. You are currently maintaining that representation in an activated state (e.g., via sustained neural firing), such that presenting the stimulus "WM" easily brings to mind the concept those letters stand for (and primes additional WM-related knowledge you may have stored in your long-term memory). Alongside "WM," you are also storing transient representations of individual words in this paragraph, most of which are briefly entering your focus of attention before being quickly replaced by new information as you progress through each sentence. If we give you additional items to hold in mind (e.g., O, +, □, ×), you will quickly find yourself overloaded and struggling to keep up with the meaning of the text. Here your WM has reached its capacity limit, resulting in impaired performance. However, it is possible to free up capacity using volitional strategies, such as mentally grouping items into singular representations (e.g., ⊕, ⊠) or removing items you no longer need. During this paragraph, you have experienced several subprocesses critical to WM, including encoding, maintenance, retrieval, updating, removal, and substitution. You have also experienced performance costs associated with WM's extremely limited storage capacity and their alleviation through a strategy called "chunking." In this section, we introduce the concept of WM and outline prevailing theoretical perspectives from cognitive psychology and neuroscience.

As you have just experienced, WM refers to a set of processes that makes task-relevant information accessible to higher-level cognitive processes, such as reading comprehension, planning, abstract reasoning, decision-making, problem solving, and learning (Baddeley, 1992; Cowan, 1988; Daneman & Carpenter, 1980; Kyllonen & Christal, 1990; Miller et al., 1960). WM has been described as a "mental workspace" (Logie, 1995) and "the sketchpad of conscious thought" (Baddeley, 1992; Miller et al., 2018). It allows information to be kept "in mind" over short timescales (Goldman-Rakic, 1995) and is central to the organization of goal-directed behavior (Chatham & Badre, 2015; Miller & Cohen, 2001). In addition, impairments of WM feature prominently in a wide variety of neuropsychiatric disorders (e.g., attention-deficit/hyperactivity disorder (Ortega et al., 2020), schizophrenia (Silver et al., 2003)). Understanding WM is thus of key interest to both cognitive and clinical neuroscientists.

In terms of its structure, prominent cognitive theories view WM as an activated subset of mnemonic information stored in either long-term memory (LTM) or short-term memory (STM) (Fig. 1) (Cantor & Engle, 1993; Cowan, 1988, 1999, 2008, 2016, 2019; Oberauer, 2002, 2009). This activated subset rises to conscious awareness and becomes available for further processing via selective attention (D'Esposito, 2007; Oberauer, 2002, 2009). However, WM differs conceptually

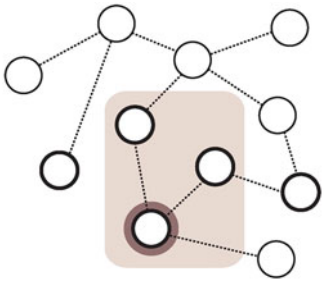

Fig. 1 Oberauer's (2002) model of working memory. Nodes and lines represent a network of mnemonic representations stored in LTM, some of which are activated (bold nodes). A subset of activated items is held (via item-context bindings) in a limited capacity *region of direct access* (shaded area) corresponding to the storage capacity of WM. Within this region, one item is selected for processing by the focus of attention (highlighted bold node). (From Bledowski et al. (2010) with permission)

from both LTM and STM (Cowan, 2008, 2017). LTM typically refers to unlimited capacity storage of information over long timescales (e.g., days, months, or years) (Cowan, 2008; Shiffrin & Atkinson, 1969). In contrast, WM is extremely capacity-limited (current work suggests a limit of between one and four representations or "items," Cowan, 2001; Garavan, 1998; Sewell et al., 2014, 2016), and its contents are quickly forgotten once attention is withdrawn (for a review, see Ricker et al. (2016)). Likewise, STM refers to passive storage (i.e., without manipulation) over short timescales (Cowan, 2017; Diamond, 2013; Engle et al., 1999), whereas WM involves actively "working on" or transforming information in a manner that supports flexible goal-directed behavior (for further distinctions, see Cowan (2008)). Much of this chapter will focus on the fundamental subprocesses of WM that carry out this cognitive "work."

The information stored in WM is typically spoken of in terms of *items* or *cognitive objects*, which denote distinct mnemonic representations that are retrieved and operated on as a unit (Mathy & Feldman, 2012; Thalmann et al., 2019). The complexity of what an item may represent can vary greatly. For example, simple perceptual stimuli (e.g., line orientations, color and shape stimuli, numbers, single words), complex multi-attribute or composite stimuli (e.g., human faces, natural scenes, mechanical objects), and more abstract and internally generated mental quantities (e.g., learned associations, the subjective value of choice options, the results of mental calculations) can all be represented in WM and subsequently compared, reasoned about, and used to guide goal-directed behavior (Cary & Carlson, 2001; Hardman & Cowan, 2015; Logie et al., 1994; Luck & Vogel, 2013; Santangelo et al., 2015). At a cellular level, single-cell recordings of prefrontal neural firing in monkeys show that items in WM are represented by distinct prefrontal neural assemblies that exhibit sustained item-specific neural firing while the item is being maintained (i.e., during the retention interval in delayed-response tasks, Fig. 2) (Funahashi et al., 1989, 1990, 1991; for a review, see Goldman-Rakic

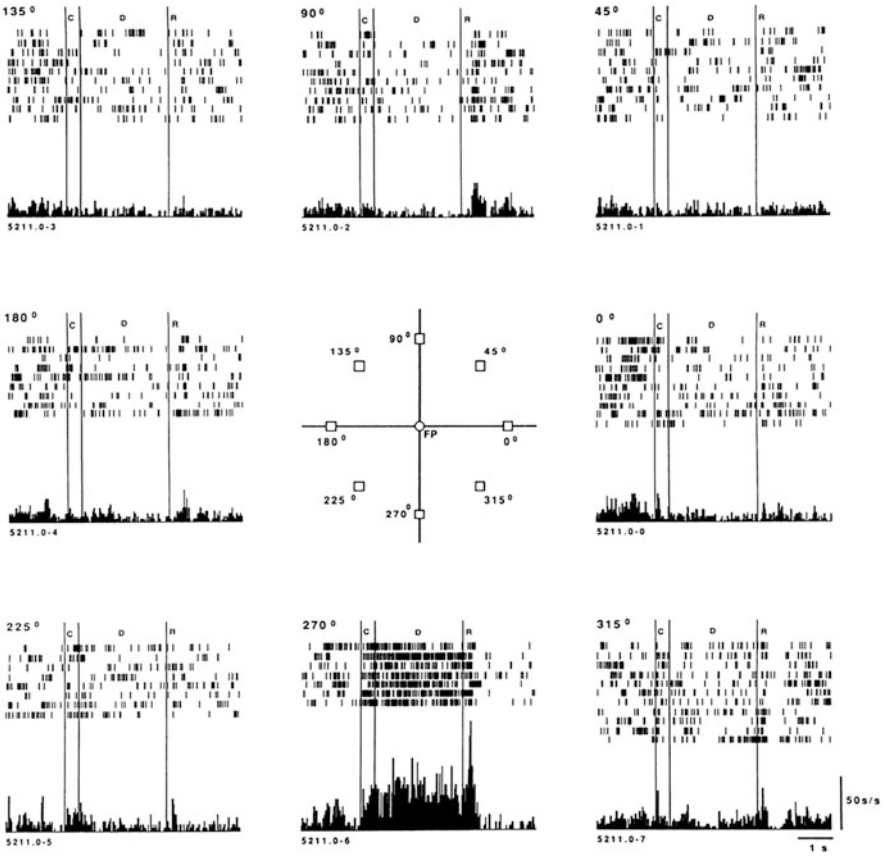

Fig. 2 A single neuron's response to eight differently oriented stimuli in a monkey performing a delayed-response WM task. The histograms depict the neuron's mean firing rate (averaged over repeated trials) during the cue (C), delay (D), and response (R) phases for each WM item. The lower middle plot shows that this neuron exhibits a sustained increase in firing during the retention interval (i.e., during maintenance) only for the item oriented 270° from the central focal point and not for the other items. WM items are thus encoded in distinct neural assemblies. (From Funahashi et al. (1989) with permission)

(1995)). Functional neuroimaging in humans has found analogous activity in the prefrontal cortex (PFC) related to maintaining distinct item representations during delayed-response tasks (for a review, see Curtis and D'Esposito (2003)).

Building on the early work with monkeys, cognitive neuroscientists have sought to identify the human brain networks involved in WM and to explain how information flows within those networks (D'Esposito, 2007). This work has used advanced neuroimaging and electrophysiological techniques – especially fMRI and EEG – to identify brain networks active during WM tasks in general, as well as neural signatures of specific operations (e.g., encoding, maintenance, updating, retrieval)

and experimental manipulations (e.g., WM load/set size, retention interval) (e.g., D'Esposito (2007); Kornblith et al. (2016); Murty et al. (2011); Riggall and Postle (2012); Roux et al. (2012)). Brain structures implicated in WM include prefrontal and parietal cortical regions, hippocampus, thalamus, basal ganglia, and dopaminergic midbrain nuclei (Bledowski et al., 2010; Bledowski et al., 2009; Durstewitz & Seamans, 2002; Murty et al., 2011; Ranganath et al., 2005). The brain's dopamine networks are thought to play a central role in controlling WM (Cools, 2019; Cools & D'Esposito, 2011; Durstewitz & Seamans, 2002). This work has given rise to a number of neurophysiologically inspired computational models that provide detailed accounts of the neural networks involved in WM and offer a coherent lens through which to interpret patterns of brain activity in WM tasks (e.g. Frank et al. (2001); Hazy et al. (2006); O'Reilly (2006); O'Reilly and Frank (2006)). However, as will be discussed, our current understanding of the neural basis of WM is far from complete.

Working in parallel, cognitive and mathematical psychologists have pursued a mechanistic understanding of various behavioral phenomena that arise in tasks involving WM (Baddeley, 1992; Cowan, 1988; Miller et al., 1960; Ratcliff, 1978). This work seeks to understand the latent cognitive processes and architectural constraints that give rise to WM-related behavioral phenomena. Much theoretical and practical interest lies in explaining *failures* of WM (e.g., unintentional forgetting) and in identifying the source of WM-related *action slips* and *performance costs* (e.g., choice/response errors, longer response latencies) that typically arise when additional demands are placed on the WM system (Oberauer et al., 2018). Typical demands include increasing the number of items to be remembered, preventing rehearsal, extending the retention interval, and adding distractors or secondary tasks. The resulting effects on behavior are considered to be fundamental or "benchmark" phenomena that good models of WM must be able to explain (Oberauer et al., 2018). The effort to understand these phenomena has generated rich theoretical debate and spawned a variety of plausible explanatory mechanisms (e.g., encoding/item fidelity, capacity limitations, temporal decay, inter-item interference, retrieval competition) that have been formalized in computational cognitive models (Farrell & Lewandowsky, 2002; Oberauer, 2009; Oberauer & Lewandowsky, 2011, 2016; Ratcliff, 1978; Sewell et al., 2016). Many of these models successfully explain a subset of the benchmark phenomena, although no model provides a unified account of them all.

Overall, this work has provided significant insight into the mechanisms and structure of WM and its relation to other cognitive systems (e.g., attention, learning, decision-making). However, relatively little attention has been given to the question of how the information in WM remains relevant to the organism's current situation and how that information is organized to best serve the organism's goals. Recently, it has been proposed that WM relies on a set of operations or subprocesses such as the gating of new incoming information, item removal and substitution, and reorganization operations (e.g., Ecker et al. (2014b), Hazy et al. (2006), Oberauer and Lewandowsky (2016), Rac-Lubashevsky and Kessler (2016b), Timm and Papenmeier (2019)). Given the foundational importance of such algorithmic

subprocesses, this chapter seeks to summarize our current state of knowledge of the cognitive and neural basis of these subprocesses and to suggest a path forward using methods that we believe have potential to provide a more complete neurocomputational understanding of WM.

This chapter is organized as follows. Having introduced the concept of WM and its theoretical background, we now discuss a central computational problem faced by the organism: the *stability-flexibility tradeoff*. Drawing on current neurocomputational theory, we explain the mechanisms by which WM is used to resolve this tradeoff and the brain networks in which they are instantiated. We then outline benchmark behavioral and neural evidence for key WM subprocesses involved in resolving this tradeoff and facilitating flexible goal-directed behavior. We draw attention to several limitations associated with common analysis methods and point to recent advances in model-based cognitive neuroscience that could further our understanding of WM subprocesses. We conclude with a discussion of several current "hot topics" in WM research which serve to place our picture of WM in the broader context of current theoretical work in cognitive neuroscience.

1.2 The Stability-Flexibility Tradeoff

As mentioned, WM is extremely capacity-limited. It is currently thought that only between one and four items can be maintained in an activated state in WM at a time (Cowan, 2001; Garavan, 1998; Oberauer, 2002; Sewell et al., 2014). This severe limit demands a high degree of control over WM content to ensure that only relevant information occupies WM (Dreisbach, 2012). To this end, WM must strike a balance between *stability* (i.e., protecting the current contents of WM from irrelevant or distracting information) and *flexibility* (i.e., keeping WM up to date with relevant new information and removing outdated information) (Dreisbach, 2012; Dreisbach & Fröber, 2019; Dreisbach et al., 2005; Hommel, 2015; Oberauer, 2009). Resolving this stability-flexibility tradeoff is a central problem for numerous executive control processes, including task switching, inhibition, and conflict monitoring/resolution (Musslick et al., 2018), and striking the right balance depends critically on the brain's dopamine systems (Cools, 2019; Cools & D'Esposito, 2011; Durstewitz & Seamans, 2008). As you experienced in the opening paragraph, managing this tradeoff is important for goal-directed behavior in dynamic environments, in which the relevance of information is continually changing, and distracting information frequently competes for attention.

To resolve the tradeoff, WM performs *information gating*, which is accomplished by switching between two different operating modes, that is, *maintenance* and *updating* (Badre, 2012; Bledowski et al., 2010; Durstewitz & Seamans, 2008; Kessler & Oberauer, 2014; Miller & Cohen, 2001; Murty et al., 2011; O'Reilly, 2006; Roth et al., 2006). The maintenance mode prevents irrelevant and distracting information from interfering with the current contents of WM, while the updating mode allows new information to enter (and old information to exit)

WM. Maintenance is supported by volitional strategies (e.g., articulatory rehearsal, refreshing) and cognitive control processes (e.g., gating, inhibition) that serve to strengthen or refresh item representations and their bindings to contextual cues (Morey & Cowan, 2018) and filter out or reduce the influence of distractors (Oberauer & Lewandowsky, 2016). Strategies that support effective remembering by restructuring information into more memorable formats (e.g., via "chunking" and sorting) also contribute to robust maintenance (D'Esposito et al., 1999; Marshuetz, 2005; Nassar et al., 2018; van Dijck et al., 2013). Likewise, updating is supported by a number of subprocesses that ensure new information can enter WM and old information can leave (Ecker et al., 2014a, b; Oberauer, 2018; Rac-Lubashevsky & Kessler, 2016b). For example, *removal* deactivates previously relevant information, freeing up capacity to activate different information (Morey & Cowan, 2018), while *substitution* replaces one item with another, allowing new information to be "loaded in" to WM (Rac-Lubashevsky & Kessler, 2016b). Together, these subprocesses allow WM to alternate modes between *flexible*, when new information is encountered, and *stable*, when information must be shielded from distractors.

In terms of the brain networks that control switching between maintenance and updating, the most prominent neurocomputational theory is the prefrontal cortex basal ganglia WM model (PBWM) (Frank et al., 2001; Hazy et al., 2006; O'Reilly & Frank, 2006). The PBWM proposes that switching between maintenance and updating modes is accomplished via "go/no-go" signaling between the basal ganglia, thalamus, and PFC. As illustrated in Fig. 3, gate opening is controlled by a striatal "go" signal[1] which inhibits substantia nigra pars reticulata (SNr) and disinhibits thalamus, which in turn excites specific neural populations of the PFC. This allows information to enter WM and updating to occur. Gate closing is controlled by a striatal "no-go" signal which inhibits external globus pallidus (GPe), disinhibits SNr, and inhibits thalamus. This in turn keeps the PFC inhibited, which prevents WM from being updated (Hazy et al., 2006). In short, the "go" signal passes through two inhibitory connections (striatum-SNr-thalamus), which excites PFC, while the "no-go" signal passes through three inhibitory connections (striatum-GPe-SNr-thalamus), which inhibits PFC. This gating network has received broad support from a number of functional magnetic resonance imaging (fMRI) studies identifying regions that are selectively active during either updating or maintenance

[1] The PBWM model suggests a phasic dopaminergic signal from the midbrain dopamine structures only in the early phases of a WM task when the BG must learn when to update. Once WM updating rules are learned, BG nuclei no longer rely on a phasic dopaminergic response but control WM gating via the non-dopaminergic SNr. Any additional dopaminergic input reflects either reward associations or a feedback-based response which evaluates the updating process based on the reward prediction error coded by the same neurons (Schultz, 1998; Schultz et al., 1997). This response, in the form of bursts and dips in dopaminergic release onto striatal neurons, is thought to reinforce "go" and "no-go" activation, respectively. In addition, the PBWM assumes that WM sits in the maintenance or "gate closed" mode by default. This assumption is likely too strong since it implies that gate opening must always accompany updating. Under this assumption, the PBWM would fail to predict the different gating costs to WM updating that occur in behavioral data (Rac-Lubashevsky & Kessler, 2016a, b).

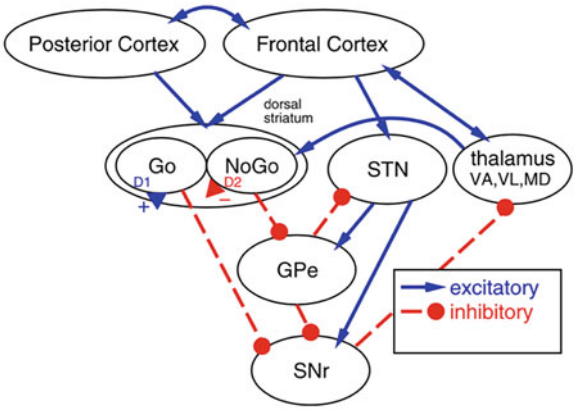

Fig. 3 Illustration of the PBWM model. Several parallel loops connect BG and the frontal cortex. Gate opening is controlled by a striatal "go" signal that inhibits SNr and disinhibits thalamus and PFC, enabling updating to occur. Gate closing is controlled by a striatal "no-go" signal that inhibits GPe, disinhibits SNr, and inhibits thalamus and PFC, preventing updating. BG nuclei learn when to perform this signaling through environmental rewards and punishments. Extending this model to include additional structures implicated in WM and cognitive control and their role in WM subprocesses beyond gate opening and closing is a key target for model-based cognitive neuroscience. (From Hazy et al. (2006) with permission)

(Cools et al., 2007b; Dahlin et al., 2008; Lewis et al., 2004; McNab & Klingberg, 2008; Murty et al., 2011; Riggall & Postle, 2012; Roth et al., 2006; Tan et al., 2007). Moreover, the same network plays a similar role in updating value representations in reinforcement learning and value-based decision making (see Chapter "Cognitive Modeling in Neuroeconomics"), suggesting that it may be a domain general neural mechanism for accomplishing information gating (Bledowski et al., 2010; Cools et al., 2007a; Hazy et al., 2006; Jocham et al., 2011; Möller & Bogacz, 2019; O'Reilly, 2006; O'Reilly & Frank, 2006; Roth et al., 2006).

The PBWM provides a coherent framework through which to understand BG and PFC activity in WM-based tasks. However, there are numerous outstanding questions regarding the neural basis of several WM subprocesses. The following subsections take a closer look at behavioral and neural evidence for key WM subprocesses, through the lens of model-based cognitive neuroscience (Corrado & Doya, 2007; Forstmann & Wagenmakers, 2015; Friston, 2009; Love, 2016; Trutti et al., 2021; Turner et al., 2017a, 2019b). In doing so, we place several benchmark behavioral phenomena in the context of current cognitive and neurocomputational theory and draw attention to methodological challenges associated with linking brain and behavior, which may be resolved through the methods of model-based cognitive neuroscience.

2 Revealing the Subprocesses of WM

WM relies on a number of subprocesses to manipulate item representations and ensure only relevant information is maintained in WM. Recent experimental work using novel WM tasks (e.g., the reference-back and multiple-item updating paradigms, Box 1) (Ecker et al., 2014a; Rac-Lubashevsky & Kessler, 2016b) has identified behavioral signatures of several WM subprocesses. These include *information gating* and *rehearsal/refreshing* processes that support maintenance and effective remembering, *item-specific removal* and *substitution* mechanisms that actually perform updating in WM, and processes that *retrieve*, *select*, and *manipulate* information when multiple items are represented in WM (Ecker et al., 2014a, b; Jongkees, 2020; Lewis-Peacock et al., 2018; Nir-Cohen et al., 2020; Rac-Lubashevsky & Kessler, 2016a, b, 2018, 2019; Rac-Lubashevsky et al., 2017; Verschooren et al., 2021). Before outlining behavioral and neural evidence for these processes, we briefly mention the features of experimental tasks used to study WM subprocesses in the laboratory.

The details of tasks used to study WM subprocesses can vary greatly (Box 1). However, they share a general structure: Participants must hold one or more items in WM over a *retention interval* (typically over a brief delay period or number of successive experimental trials), after which their memory for those items is tested. On each trial, participants are cued to maintain, update, remove, or otherwise manipulate one or more of the items in WM before responding to a memory probe via button-press response. Probe stimuli may be targets or distractors, which match or mismatch the contents of WM, respectively. Responses thus require a decision based on WM representations, the quality of which affects the difficulty of decisions and has systematic effects on errors and response latencies (Donkin et al., 2016; Ratcliff, 1978; Sewell et al., 2016). Performance is often analyzed using Donders' (1969) *subtraction method*, whereby behavioral measures (e.g., error rate, mean response time) are compared between trial types (e.g., maintain vs. update, high vs. low load) in order to isolate the behavioral costs unique to the WM subprocess of interest. Likewise, comparing neurophysiological measures between trial types (e.g., via event-based fMRI and EEG) isolates neural activity unique to particular subprocesses (e.g., fMRI BOLD activity unique to updating, EEG theta power unique to maintenance; Murty et al., 2011; Rac-Lubashevsky & Kessler, 2018). As we will see, a current priority of model-based cognitive neuroscience is to develop formal models of WM that explain behavioral measures in terms of latent cognitive processes and link these to brain data through joint and integrated "brain-cognition-behavior" models (Trutti et al., 2021) (see Chapters "Linking Models with Brain Measures" and "Advancements in Joint Modeling of Neural and Behavioral Data").

Box 1 Measuring WM Subprocesses with the Reference-Back and Multiple-Item Updating Paradigms

The Reference-Back Task

Most laboratory tasks used to study WM (e.g., *n*-back, delayed-match-to-sample) are designed to investigate the capacity and temporal properties of WM but are unable to differentiate the contribution of WM subprocesses to observed behavior (Ecker et al., 2010; Lewis-Peacock et al., 2018; Nir-Cohen et al., 2020; Rac-Lubashevsky & Kessler, 2016a, b; Roth et al., 2006). For example, in the widely used *n*-back task, in which the subject updates the serial position of items in WM on each trial, several cognitive processes occur together on every trial and thus cannot be dissociated. These include encoding the incoming stimulus and binding it to a position in WM, comparing the incoming stimulus with the correct reference item, inhibiting items in irrelevant positions (i.e., avoiding intrusion effects), updating the position of each item in WM, and removing items that are no longer relevant. A recently developed exception is the *reference-back* task (Rac-Lubashevsky & Kessler, 2016a, b), which provides dissociable measures of core WM subprocesses (gate opening, gate closing, updating, substitution) from behavioral (choice-response time) data. The task is illustrated in Fig. 4.

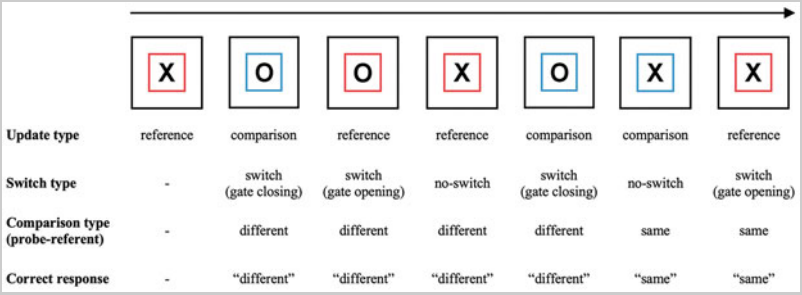

Fig. 4 Illustration of the reference-back task. On each trial, participants indicate whether the presented letter is same or different from the letter in the most recent red frame. On reference (red frame) trials, participants must also update WM with the currently displayed letter. On comparison (blue frame) trials, participants make the same/different decision but do not update WM. (Adapted from Rac-Lubashevsky and Kessler (2016b) with permission)

To perform the reference-back, participants hold one of two stimuli (e.g., an "X" or "O") in WM while deciding whether a series of probes match the current item in WM. On *reference trials* (indicated by a red frame around the stimulus), the participant must update WM with the currently displayed stimulus. On *comparison trials* (indicated by a blue frame), the participant simply compares the current stimulus to the one held in WM (the one

(continued)

appearing in the most recent red frame) without updating WM. Both reference and comparison trials require a same/different decision but only reference trials require updating. Comparing performance on reference and comparison trials thus provides a behavioral measure of the cost of *updating*.

By similar logic, switching from comparison to reference trials requires opening the WM gate (to allow for updating), while switching from reference to comparison trials requires closing the WM gate (to maintain the current contents). *Gate opening* is measured by comparing trials on which participants switch toward a reference trial to those where reference trials are repeated. Likewise, *gate closing* is measured by comparing trials on which participants switch toward a comparison trial to those where comparison trials are repeated. Finally, *substitution* is measured via the interaction effect of trial type (reference/comparison) and match type (same/different) and represents the cost of updating a *new* item into WM.

The benchmark behavioral finding from the reference-back task is that trials requiring additional WM processes tend to have slower RTs and/or more frequent errors than trials that do not require such processes (Jongkees, 2020; Lewis-Peacock et al., 2018; Nir-Cohen et al., 2020; Rac-Lubashevsky & Kessler, 2016a, b, 2018, 2019; Rac-Lubashevsky et al., 2017; Verschooren et al., 2021). In terms of latent cognitive processes, these costs are typically interpreted as reflecting a combination of time required for additional subprocesses to run outside of the same/different decision stage and subprocesses interfering with the primary task (e.g., creating noisier WM representations due to drawing attention/capacity away from the decision process) (Pearson et al., 2014). Decomposing behavioral costs on the reference-back into latent cognitive processes and linking them to neural activity is a current goal of model-based cognitive neuroscience (for a review, see Trutti et al. (2021)).

The Multiple-Item Updating Task

A close relative of the reference-back is the multiple-item updating paradigm (Ecker et al., 2014a, b), which has been used to distinguish between item-specific and global removal processes in behavioral (choice-response time) data.

In this task, participants are presented with a row of three stimuli (in black frames) to hold in WM. Participants then perform an unpredictable number of updating steps on which one or more items must be updated with a new stimulus. Crucially, the to-be-updated positions are sometimes pre-cued (indicated by an empty red frame), which is assumed to give the participant time to remove the corresponding WM item prior to encoding its replacement. Pre-cueing a position for removal typically results in performance benefits (e.g., in updating speed) for the cued items, which suggests the existence of an

(continued)

item-specific removal process. Cueing participants to remove one, two, or all three items in this paradigm has also allowed researchers to make distinctions between item-specific and global removal processes (see main text for further details).

2.1 Closing the Gate (Entering Maintenance Mode)

The first subprocess we consider is *gate closing*, which puts WM into maintenance mode via the BG-PFC signaling mechanism described earlier. In this mode, the contents of WM are shielded from distracting information, which allows the organism to maintain stable item representations (Verschooren et al., 2021; Verschooren et al., 2020). Several mechanisms contribute to stability, including those that counteract temporal degradation by boosting the activation of relevant information (e.g., articulatory rehearsal, refreshing) and those that reduce interference by diminishing the influence of irrelevant information (e.g., filtering, inhibition) (Morey & Cowan, 2018; Oberauer & Lewandowsky, 2016). Based on functional anatomy of the BG (Parent & Hazrati, 1995), it is typically assumed that WM sits in the gate-closed/maintenance mode by default (e.g., Frank et al. (2001), Hazy et al. (2006)) since this strategy frees the organism from having to continuously exert effort to maintain WM items in the face of distractors. The processing that occurs in the maintenance mode is also thought to facilitate the formation and consolidation of long-term memories, mediated by a network that includes hippocampus and dorsolateral PFC (Ranganath et al., 2005).

Behaviorally, gate closing occurs when switching from a trial requiring updating to a trial requiring maintenance. Comparing performance to non-switch trials (i.e., repeated maintenance trials) measures the cost of closing the gate. The benchmark finding is that performance on trials that require closing the gate is slower and more error prone than on trials that do not require closing the gate (Kessler & Oberauer, 2014, 2015; Rac-Lubashevsky & Kessler, 2016a, b). It is currently unknown whether these gating costs are due to time added outside of the decision-making stage or interference with the primary task (e.g., by drawing resources such as attention away from the decision process) (Pearson et al., 2014). As we discuss below, formal cognitive modeling that decomposes gating costs into latent cognitive processes is needed to answer such questions.

Neuroscientific work has primarily used event-based contrasts of fMRI and EEG signals (e.g., on maintenance vs. updating trials) to identify statistically significant differences in neural activity between updating and maintenance modes that can be attributed to gating. Several cortical areas previously related to cognitive control, such as dorsolateral PFC, medial PFC, and parietal cortex, exhibit BOLD activity uniquely associated with maintenance and not updating (D'Esposito & Postle, 2015;

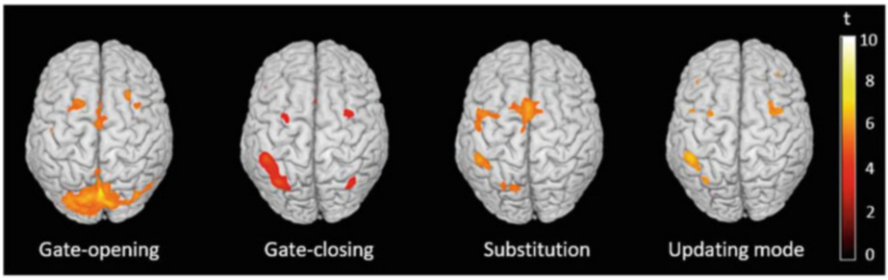

Fig. 5 Whole-brain analysis of neural substrates of several WM subprocesses as measured by the reference-back task (Rac-Lubashevsky & Kessler, 2016a). Red/yellow regions represent statistically significant contrasts in fMRI BOLD activation between reference-back trial types (see Box 1 for details). (From Nir-Cohen et al. (2020) with permission)

Feredoes et al., 2011; Nir-Cohen et al., 2020; Owen et al., 2005; Roth et al., 2006). Gate closing has been linked to increased theta power in EEG oscillatory signals (Rac-Lubashevsky & Kessler, 2018), which is also a widely reported neural signature of cognitive control (Cavanagh & Frank, 2014; Cohen, 2014; Cohen & Donner, 2013), while gamma-band oscillations have been linked to the active maintenance of WM information (Roux & Uhlhaas, 2014). Dorsolateral PFC and parietal cortex – particularly the superior intraparietal sulcus – are active when maintaining stable representations in the presence of distractors (Behrmann et al., 2004; Bettencourt & Xu, 2015; Clapp et al., 2010; Dolcos et al., 2007; Durstewitz et al., 2000; Lorenc et al., 2018; Sakai et al., 2002; Toepper et al., 2010), and the insula has been implicated in sustaining attention during maintenance (Dosenbach et al., 2008). Figure 5 shows some examples of fMRI BOLD activity related to WM processes, including gate closing, in a recent study using reference-back paradigm (see Box 1).

2.2 Opening the Gate (Entering Updating Mode)

Gate opening is arguably the most widely studied of the WM subprocesses and is critical for flexible goal-directed behavior. Opening the gate allows other WM subprocesses, such as substitution and removal (discussed below), to alter the contents of WM (e.g., by removing and replacing outdated information). As outlined earlier, neurocomputational theory suggests that gate opening is controlled by striatal "go" signaling in BG-thalamus-PFC networks (Frank et al., 2001; Hazy et al., 2006; O'Reilly & Frank, 2006) in which BG nuclei learn to signal when to update items in response to updating cues or the expectation of reward.

The behavioral costs of opening the gate parallel those of gate closing: Performance on trials that require the gate to open (e.g., when switching from maintenance to updating trials) is typically slower and more error prone than on trials that do not

require the gate to open (e.g., when performing repeated updating trials) (Kessler & Oberauer, 2014, 2015; Rac-Lubashevsky & Kessler, 2016a, b). In addition, gate opening and closing costs are often asymmetrical, with gate closing reportedly taking longer than gate opening (Ecker et al., 2014b; Rac-Lubashevsky & Kessler, 2016b). This may reflect additional effort required to reacquire representational stability versus "releasing" WM from the controlled maintenance state (Rac-Lubashevsky & Kessler, 2016a, b). Developing computational models that predict asymmetrical gating costs from a set of cognitively plausible mechanisms will be a key goal for the model-based cognitive neuroscience of WM.

Neuroscientific work largely supports the involvement of BG-thalamus-PFC networks in opening the gate to WM: striatal and dorsolateral PFC involvement has been reported in tasks broadly requiring mode switching and/or updating of WM (Dahlin et al., 2008; Lewis et al., 2004; McNab & Klingberg, 2008; Tan et al., 2007). Several studies have reported activity in subcortical structures (e.g., substantia nigra, ventral tegmental area, caudate) and frontoparietal cortical regions specific to updating and not maintenance (Bledowski et al., 2009; Lepsien et al., 2005; Murty et al., 2011; Roth et al., 2006). A recent fMRI study of the reference-back task found activity unique to gate opening in BG and frontoparietal cortex, as well as task-relevant sensory areas such as visual cortex (Nir-Cohen et al., 2020), which may be involved in encoding new information during updating (Roth et al., 2006). Striatal dopamine receptor-expressing neurons and dopamine-producing midbrain structures have also been implicated in WM updating (Cools et al., 2007b; McNab & Klingberg, 2008; Murty et al., 2011), and dynamic causal modeling suggests that BG plays a central role in gating information to PFC (van Schouwenburg et al., 2010). Further indirect support for striatal dopamine involvement comes from a study linking event-based eyeblink rate (a proxy measure of striatal dopamine) to WM updating in the reference-back task (Rac-Lubashevsky et al., 2017).

Electrophysiological correlates of gate opening and updating have also been found using EEG. Some examples are shown in Fig. 6. Delta power in EEG oscillatory signals has been associated with gate opening and updating in the reference-back task (Rac-Lubashevsky & Kessler, 2019). Delta power is a common neural signature of reactive or "event-driven" control and action selection processes (Cavanagh, 2015; Gulbinaite et al., 2014; Harmony, 2013) (see Chapter "An Introduction to EEG/MEG for Model-Based Cognitive Neuroscience"). The same study found that anterior cortical regions showed an enhanced negative ERP component that was unique to updating (Rac-Lubashevsky & Kessler, 2019). This latter signal is believed to be involved in controlled inhibition and action selection (Folstein & Van Petten, 2008) and may reflect a gate opening or updating signal consistent with the theorized role of the striatum in BG-thalamus-PFC signaling. Overall, the brain networks that control gate opening/updating in WM have clear links to the striatal reward prediction error signaling observed in reinforcement learning and value-based decision-making, in which the striatum signals when to update item representations in response to feedback from the environment (Schultz, 1998; Schultz et al., 1997). Indeed, recent work has begun to build integrated models of WM and reinforcement learning, in which behavior is the result of a mixture of

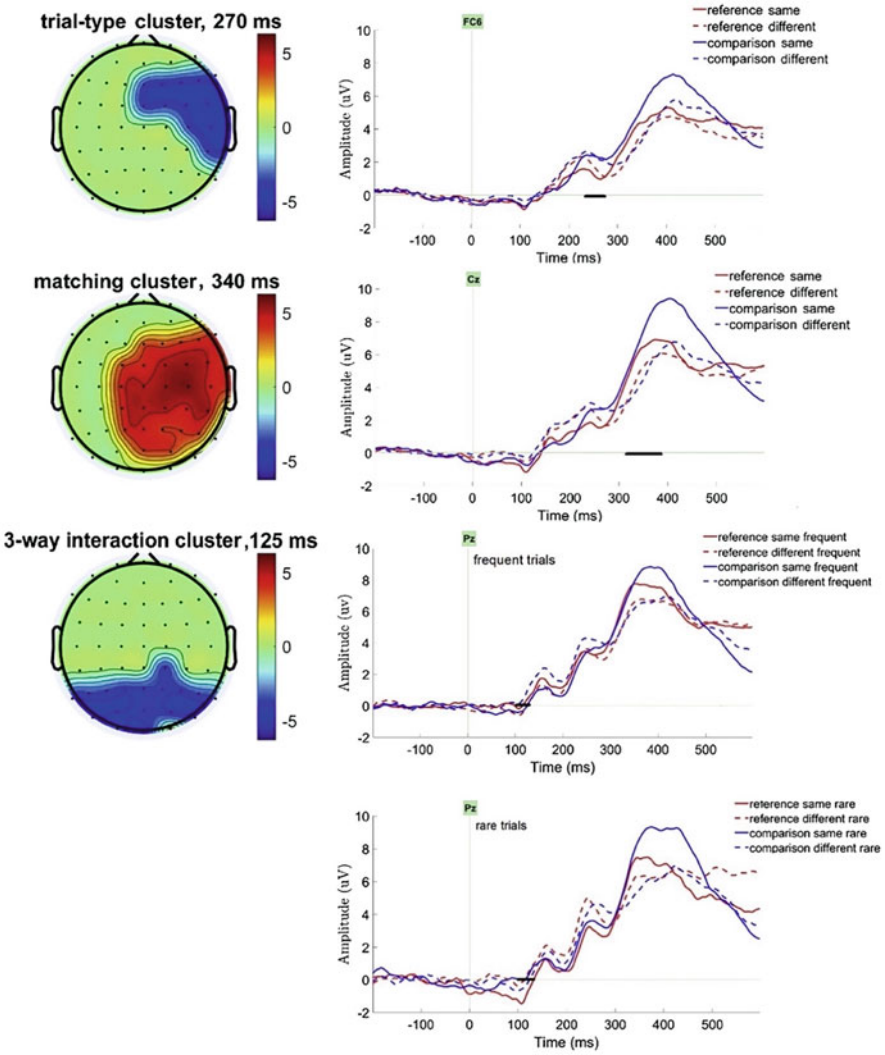

Fig. 6 EEG analysis of WM subprocesses as measured by the reference-back task. The left panel shows scalp map topography for trial type (updating), matching, and an interaction effect involving stimulus frequency. The color represents the *t*-value of the mean regression weights for each electrode at the times shown. The right panel show average ERP signals from representative electrodes within the clusters on the left. (From Rac-Lubashevsky and Kessler (2019) with permission)

the two systems (e.g., Collins and Frank (2012), McDougle and Collins (2021)). This further highlights the close connection between WM and learning, and we pick up this point again when discussing current directions below.

2.3 Removing Information from WM

Forgetting is a critical aspect of WM. Due to its severely limited storage capacity, WM must be able to create storage space by removing information that is no longer relevant to the task at hand. This is thought to be facilitated by an active removal process akin to *directed forgetting* (e.g., Festini and Reuter-Lorenz (2013, 2014)), which can be deliberately engaged to remove information from WM (Ecker et al., 2014a, b; Lewis-Peacock et al., 2018; Ma et al., 2014). Removing irrelevant information from WM is crucial for goal-directed behavior because it reduces the capacity demands on WM and reduces the likelihood of retrieving false or outdated information (e.g., by eliminating potential sources of proactive interference and reallocating cognitive resources to the remaining items, which enhances their accessibility) (Festini & Reuter-Lorenz, 2013, 2014). Removal can be temporary or permanent: items can be temporarily deactivated by withdrawing attention or WM capacity (and subsequently refreshed or reactivated) or permanently removed by unbinding item-context associations from WM (Lewis-Peacock et al., 2018). It has been suggested that formal mechanisms such as Hebbian anti-learning and resetting weights on associative connections (e.g., between item-context bindings) may be involved in deactivating or reducing the accessibility of item representations (e.g., Ecker et al. (2014a)). At the neural level, removal may work by transforming items from activation-based storage to weight-based (synaptic) storage (O'Reilly & Munakata, 2000), which echoes the difference in how storage is accomplished in short- versus long-term memory (i.e., sustained neural activity vs. synaptic consolidation/long-term potentiation). Thus, removal may involve a transformation between WM and LTM.

Behavioral signatures of removal are evident in WM tasks in which participants are cued to either maintain the current contents of WM or remove one or more items before responding to a memory probe. A recent study compared trials on which participants removed either one, two, or three items from a three-item WM set (Ecker et al., 2014b). Removing all three items (i.e., "wiping the global workspace") required less time than removing a subset of one or two target items (i.e., item-specific removal). This suggests two removal modes: In *global removal*, which is rapid but indiscriminate, all items are deactivated in parallel but potentially useful information is discarded, which causes retrieval failures and forgetting. In *item-specific removal*, items are selected and deactivated in a serial manner. This is slower and more effortful than global removal but has the advantage of allowing still-relevant items to remain in an active and accessible state (Ecker et al., 2014b; Kessler & Meiran, 2008). The benefits of removal are also evident in behavior: Performance decrements due to set-size/capacity-sharing effects are alleviated following removal (Lewis-Peacock et al., 2018; Ma et al., 2014). Conversely, access to removed information is reduced following removal (resulting in slower, more error-prone retrieval), while access to remaining items is enhanced (resulting in faster, less error-prone retrieval). These findings support the idea that removal

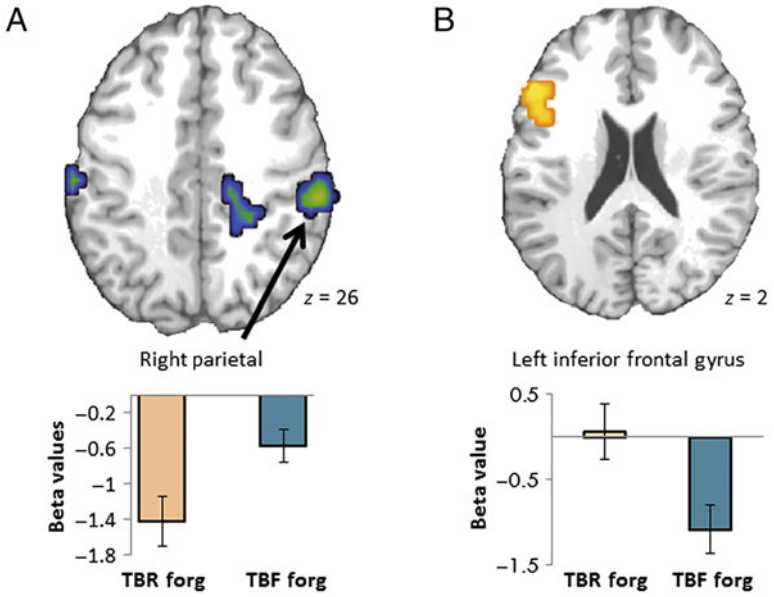

Fig. 7 Neural correlates of intentional and unintentional forgetting. In panel (**a**), the right parietal cortex showed heightened BOLD activity during intentional versus unintentional forgetting. In panel (**b**), left inferior frontal gyrus showed heightened BOLD activity during unintentional versus intentional forgetting. (From Rizio and Dennis (2013) with permission)

reduces the capacity demands on WM by making more cognitive resources (e.g., attention, WM capacity) available for processing remaining items.

Consistent with these behavioral signatures of removal, neuroscientific work has identified a number of neural phenomena related to the behavioral effects. Several studies have demonstrated that removed items have reduced neural traces at the neurophysiological level, such that probing a previously removed item elicits a muted neural response compared with other active items or the same item before removal (e.g., Lewis-Peacock et al. (2012)). Work using fMRI has reported activity in areas related to cognitive control and inhibition (e.g., prefrontal and parietal cortex) uniquely related to intentional removal as opposed to unintentional forgetting (Fig. 7) (Nowicka et al., 2011; Reber et al., 2002; Rizio & Dennis, 2013; Wylie et al., 2008). Work with EEG has identified sustained positive activation in dorsolateral PFC in response to forgetting cues (Hauswald et al., 2011; Paz-Caballero et al., 2004). Overall, this is consistent with a prefrontal network involved in inhibitory control over the contents of WM (see also Anderson and Hanslmayr (2014), Lewis-Peacock et al. (2018), and Ma et al. (2014)). One goal of future research will be to delve deeper into the difference between global and item-specific removal and establish whether distinct neural substrates underlie each process.

2.4 Substituting Items in WM

Closely related to removal is *substitution*. Substitution occurs when an item in WM is replaced by an incoming item. This allows updating to occur when WM is already occupied by active representations. In the laboratory, researchers typically study substitution by comparing the relative cost of performing different kinds of substitutions, such as overwriting active items versus empty slots (Ecker et al., 2014a) or substituting an item for an identical item (i.e., refreshing) versus a different item (i.e., replacement) (Rac-Lubashevsky & Kessler, 2016a, b). Overwriting active representations typically takes longer than overwriting empty slots (e.g., following removal) (Ecker et al., 2014a). Similarly, refreshing the same representation is less costly than replacing it with a different one (Rac-Lubashevsky & Kessler, 2016a, b). This is because replacing active information presumably requires first running the removal process, which is not needed when writing to an empty position or refreshing an existing representation.

The neural basis of substitution has received less attention than the other subprocesses. However, it likely involves many of the same dopaminergic midbrain and prefrontal cortical regions as occurred in updating more broadly (e.g., Murty et al. (2011)). A recent fMRI investigation of the reference-back paradigm found that substitution elicited unique activation in the left dorsolateral PFC and inferior parietal lobule (Fig. 5) (Nir-Cohen et al., 2020). A whole-brain conjunction analysis revealed shared activity in the supplementary motor area related to updating and substitution. These structures have previously been implicated in cognitive control and the control of motor movement (Bogacz et al., 2010; Forstmann et al., 2008). However, further work is necessary to fully characterize the neural substrates unique to the substitution subprocess rather than the updating mode more broadly. The connection to the motor system is particularly interesting in terms of broader theoretical perspectives in evolutionary psychology and embodied cognition, which draw links between the mental manipulation of cognitive representations and motor manipulation of physical objects in the environment (e.g., Anderson (2014), Wilson (2002)). This is likely to be an interesting and theoretically rich area of future research.

2.5 Retrieving, Selecting, and Operating on Multiple Items

When multiple items are maintained in WM simultaneously, a key problem that arises is how to single out a particular item from the set for retrieval or further processing. Processes that facilitate multiple-item WM include mechanisms of retrieval and selective attention (e.g., Cowan (1988), Ma et al. (2014), Sewell et al. (2016)), as well as grouping and reorganization operations (e.g., chunking and sorting) which aid retrieval by restructuring representational information into more memorable formats (Brady et al., 2009; D'Esposito et al., 1999; Marshuetz, 2005;

Nassar et al., 2018; Thalmann et al., 2019; van Dijck et al., 2013). These processes all involve an interplay between WM and attention such that various operations can be carried out selectively on some items and not others: Goal-directed attention to the external environment guides what enters WM and the fidelity with which it is represented. Conversely, goals and contextual information stored in WM influence what is attended to and guide action selection (Oberauer, 2019). Retrieval and selection processes also play a role in item-specific removal and substitution, since selectively removing or replacing one item from among many entails some ability to identify and select the target item. Current cognitive theory suggests that within WM, processing is further limited by a *single-item focus* of attention that can hold only one item at a time for further processing and which therefore imposes a serial processing constraint on WM operations (Oberauer, 2002, 2009).

Behavioral evidence for the single-item focus comes in the form of the *item-repetition benefit* (or alternatively, *item-switch cost*), whereby after an item in WM has been selected, selecting the same item for an immediately subsequent operation is faster than selecting a different item (Garavan, 1998; Gehring et al., 2003). This suggests that the object of a cognitive operation remains in the single-item focus of attention after the operation has been completed and thus does not need to be selected again when the subsequent operation requires the same object (Garavan, 1998; Oberauer, 2003). The item-repetition benefit has been observed across modalities (e.g., verbal WM (Garavan, 1998; Oberauer, 2003), spatial WM (Hedge & Leonards, 2013; Kübler et al., 2003)) and for several kinds of operations on the selected item (e.g., arithmetic (Garavan, 1998; Oberauer, 2003), updating (Oberauer, 2003), recognition (Oberauer, 2006)), supporting the idea of a general attentional bottleneck for accessing items in WM.

Another issue that arises with multiple items is that of WM's extremely limited storage capacity. Studies looking at capacity effects assess task performance at several WM set sizes (usually with at least one under- and one overcapacity condition, for example, a three-item condition and a six-item condition; Rademaker et al., 2012; Zhang & Luck, 2008), which gives insight into how different WM subprocesses respond (or fail) when the capacity of WM is exceeded. One benchmark finding is that selection and operating costs scale with set size (Oberauer et al., 2018), meaning that processes that operate on WM become slower and more error prone as WM load increases. Both fMRI and EEG measures yield signals that scale with the number of items in WM (Fig. 8) (Cowan et al., 2011; Howard et al., 2003; Kornblith et al., 2016; Leung et al., 2004; Manoach et al., 1997; Roux et al., 2012; Todd & Marois, 2004; Veltman et al., 2003; Vogel & Machizawa, 2004), while some work has identified signals more strongly related to the precision (Zhao et al., 2020) or complexity (Xu & Chun, 2006) of individual items rather than overall WM load.

One explanation for set size effects is framed in terms of resource sharing: Distributing a limited resource among multiple items reduces the precision or fidelity with which each item can be represented (Bays et al., 2009; Ratcliff, 1978; Sewell et al., 2016). In support of this account, participants can be directed to treat a subset of items as more important, which produces a robust gain in recall precision in emphasized items. Moreover, this gain comes at a cost to non-emphasized items,

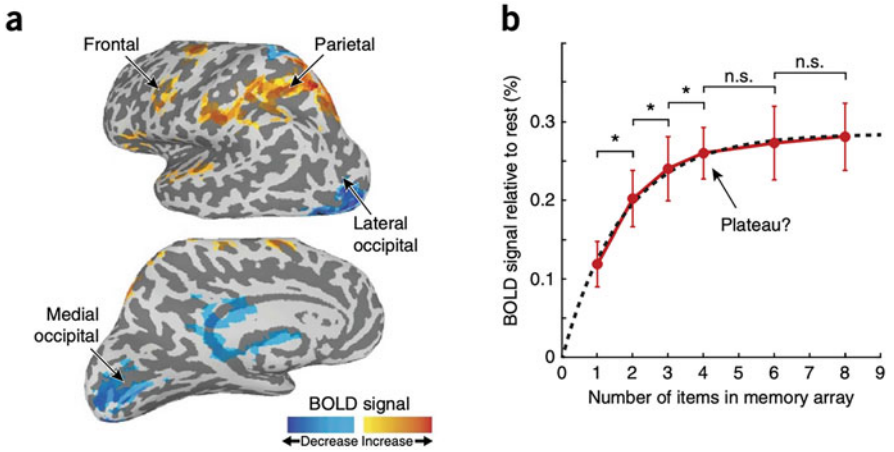

Fig. 8 Neural activity related to WM storage and capacity. In panel (**a**), short-term maintenance of visual information produces a sustained increase in BOLD activity (red and yellow) in prefrontal and posterior parietal areas and a decrease in occipital areas (Riggall & Postle, 2012). In panel (**b**), change in BOLD signal in parietal cortex scales in magnitude with the number of items being stored in WM (Todd & Marois, 2004). (Adapted from Ma et al. (2014) with permission)

which supports the idea that items are allocated resources from a limited capacity continuous pool controlled by selective attention (Bays & Husain, 2008; Ma et al., 2014). A related idea is that capacity effects arise due to retrieval competition, such that the probability of retrieving a particular target item decreases in proportion to the number of competing items (Oberauer, 2009). The retrieval competition account is also supported by the finding that the severity of inter-item interference scales with inter-item similarity (Oberauer & Bialkova, 2011). There is ongoing debate concerning whether items are allocated capacity from a limited pool of continuous resources (Bays & Husain, 2008; Ma et al., 2014; Zhang & Luck, 2008) or held in a small number of discrete high-precision slots (or some combination of the two) (Donkin et al., 2016; Donkin et al., 2013).

Finally, capacity demands can be alleviated to a certain extent by grouping and reorganization operations that restructure information into more memorable or accessible formats. Grouping or "chunking" involves binding together previously separate items or combining features to form a single representation that can be retrieved as a unit. For example, Luck and Vogel (1997) found no performance costs when comparing memory for single- versus multiple-feature items, suggesting that multiple-feature items were processed with the same efficiency as single-feature items (Luck & Vogel, 1997; but see Hardman and Cowan (2015), who found effects of both feature and item load). Reorganization involves volitional mnemonic strategies, such as sorting items into alphabetical or chronological order or otherwise organizing the contents of WM in a manner that improves recall (D'Esposito et al., 1999; Marshuetz, 2005; Nassar et al., 2018; van Dijck et al., 2013). In terms of

neural correlates, "chunking" and similar encoding strategies have been linked to prefrontal and parietal networks involved in associative learning (Bor et al., 2004; Bouchacourt et al., 2020; Pang et al., 2019).

3 Toward a Model-Based Cognitive Neuroscience of Working Memory Subprocesses

It is clear from the preceding discussion that our current picture of WM is a complex one that we are yet to fully understand. WM is supported by numerous fundamental subprocesses that perform a variety of distinct but complimentary functions that often interact in complex ways. Significant progress has been made in characterizing the behavioral signatures of various WM subprocesses, and recent work has begun to make connections to their neural substrates. Behavioral measures of WM subprocesses are typically computed using the logic of the subtraction method (i.e., taking differences between trial types; Donders, 1969) and summarized using aggregate measures of task performance such as mean response time and error rate. These are then subject to traditional statistical analyses such as ANOVA and regression. However, an important limitation of the subtraction method is that it is not a pure measure of the latent processes that give rise to the aggregate behavioral difference. Identifying the latent processes responsible requires detailed cognitive modeling of choice-response time distributions (Ratcliff, 1978). The pitfalls of using coarse behavioral measures to draw inferences about latent cognitive processes have been discussed extensively in the mathematical psychology literature (e.g., Ratcliff (1978), Wagenmakers et al. (2007)). However, model-based analyses of WM subprocesses remain exceedingly rare.

This issue is particularly important in WM, in which a large number of mechanisms operate in concert over extremely brief timescales (e.g., 10–50 ms). In practice, this means that there is a limit on how informative behavioral data (e.g., choice-response time) can be for selecting between mechanisms because the latent cognitive processes of interest tend to mimic each other at the level of behavior (Hawkins et al., 2017). This issue will likely be alleviated to some extent by the development of more detailed computational models of WM that decompose observed behavior into latent component processes or integrate WM subprocesses into models of higher-level cognition, such as reinforcement learning (e.g., Collins and Frank (2012), McDougle and Collins (2021)) and evidence accumulation models of decision-making (e.g., Brown and Heathcote (2008); Forstmann et al. (2016); Ratcliff (1978)), or more general cognitive architectures, such as ACT-R (Anderson & Lebiere, 2014; Anderson et al., 1996). Evidence accumulation models, which explain choices and response time distributions in terms of latent cognitive processes, are particularly well suited to reveal whether the WM-related phenomena outlined above occur because WM subprocesses add time outside of the decision stage (longer nondecision time), interfere with the decision process itself (reduced

or noisier processing rate), or induce strategic adjustments engaging top-down cognitive control (increased response caution). Decomposing WM phenomena (e.g., gating, switching, updating costs) into a set of latent cognitive processes rather than relying on aggregate summary statistics (e.g., mean response time) enables exploring WM subprocesses in greater detail than is typically achieved.

Neuroscience has made impressive progress in identifying unique neural signatures of WM subprocesses that are detectable in fMRI and EEG data. As we have seen, typical analysis methods involve taking event-based contrasts of neurophysiological signals (e.g., the difference in BOLD signal between maintenance and updating trials), which are sometimes regressed on or correlated with behavioral measures. The high spatial resolution of fMRI has proven particularly useful in detecting activity in subcortical structures involved in WM such as the striatum. However, most imaging of WM subprocesses has been conducted at a field strength of 3 Tesla, which offers relatively poor resolution of small subcortical structures (e.g., substantia nigra, ventral tegmental area, subthalamic nucleus) compared with ultrahigh field (e.g., 7 T and higher field strengths) machines (de Hollander et al., 2017; Miletić et al., 2020; Trutti et al., 2019). Ultrahigh field fMRI (with increased resolution and improved signal- and contrast-to-noise ratios) will likely play a significant role in identifying the neural substrates of WM with sufficient spatial precision.

Despite advances in spatial resolution, fMRI still relies upon slow changes in BOLD response following neural activity, which gives fMRI poor temporal resolution. Due to the inherently slow BOLD response, even ultrahigh field fMRI is ill-equipped to detect activity stemming from extremely brief WM subprocess that operate at timescales on the order of 1050 ms. This is particularly so if the structures that generate such brief pulses of activity are located deep within the subcortex. High temporal resolution methods like EEG, which offer a practically continuous-time read-out of neural activity, are likely to play a complimentary role in detecting neural signatures of extremely brief subprocesses and resolving the spatiotemporal tradeoff in neurophysiological measurement. However, the low spatial resolution of EEG means that it cannot unambiguously attribute electrical activity recorded from the scalp to deep subcortical sources (this requires solving the *inverse problem*, Grech et al., 2008). As discussed elsewhere in this book (see Chapter "An Introduction to EEG/MEG for Model-Based Cognitive Neuroscience"), one reason for this is that dendrites of subcortical neurons are not as regularly aligned as cortical pyramidal cells. This means that the polarity of electrical signals from subcortical structures tends to cancel out when summing over many neurons and thus signal from subcortical structures does not reach the scalp. A combination of fMRI and EEG techniques, along with experimental designs that elicit strong neural contrasts between conditions, will likely be required to achieve a complete understanding of the neural basis of WM.

From the perspective of model-based cognitive neuroscience, one of the most important limitations is that common approaches in the literature on WM subprocesses analyze brain and behavior separately, without attempting to link the two sources of data. Although conducting separate analyses is relatively simple,

such approaches ignore the mutual constraint afforded by linking behavioral and neural data in a joint analysis or integrated neurocognitive model (Turner et al., 2013, 2017a, b, 2019a, b). In cognitive neuroscience, joint modeling involves constructing a model of the behavioral data and connecting it to a model of the neural data via a linking function (see Chapter "Linking Models with Brain Measures"). The choice of behavioral model typically depends on the experimental task (e.g., speeded decision-making, reinforcement learning) and describes the latent cognitive processes that give rise to behavior. The choice of neural model typically depends on the modality of the neural data (e.g., fMRI, EEG) and describes the shape of the neural response (e.g., hemodynamic and electrophysiological response functions). Parameters and mechanisms in each model are then linked, either directly, by parameterizing one model in terms of the other, or indirectly, by imposing a hierarchical structure in which parameters are related through a common overarching distribution (e.g., multivariate normal) (see Turner et al.'s discussion of the *covariance* approach). This allows for simultaneous fitting and parameter estimation for both models that properly accounts for the measurement noise in each data source (see Chapter "Cognitive Models as a Tool to Link Decision Behavior With EEG Signals"). Joint analyses impose stronger constraints on theory and ultimately produce more detailed and reliable inferences about the latent processes that generate behavior. Analyses that do not account for mutual constraint of behavioral and neural data risk overfitting each data source and thus misattributing (or exaggerating) effects, which results in biased inferences that do not generalize well to new data.

Joint modeling also allows closer contact to be made between cognitive models and biologically plausible neurocomputational models such as the PBWM (Frank et al., 2001; Hazy et al., 2006; O'Reilly & Frank, 2006). This would greatly improve our understanding of the relation between cognitive and neural mechanisms underlying WM. For example, joint approaches can alleviate issues of model mimicry (i.e., different models making the same prediction), since the additional structure in the neural data can be used to select between models that differ in their internal dynamics but make identical predictions at the level of choice and mean response time (Ditterich, 2010; Forstmann et al., 2011; Hawkins et al., 2017; Mack et al., 2013; Purcell et al., 2010; Purcell & Palmeri, 2017; Schall, 2019). Joint modeling can also be used to "fuse" together additional data sources (e.g., behavior + fMRI + EEG; Turner et al., 2016) into a single model, the benefit of which is to improve the model's predictive accuracy and generalizability. For example, through several cross-validation analyses, Turner et al. (2016) showed that including both EEG and fMRI data in a joint model allowed for better predictions of behavioral data than did either modality alone, suggesting that the EEG and fMRI data contained slightly different information about the decision-making process. Overall, such approaches hold great promise for improving our understanding of WM subprocesses and bringing us closer to the ideal of an integrated model of WM that generates both behavioral and neural data from a common set of latent mechanisms.

We are confident that the methods of model-based cognitive neuroscience outlined above hold great promise for furthering our understanding of WM subprocesses and their neural substrates. However, some caution is warranted in interpreting joint analyses. In terms of interpreting correlations between cognitive and neural mechanisms, it is important to keep in mind that numerous neurocognitive systems (e.g., WM, S/LTM, learning, cognitive control) operate in parallel even in very simple laboratory tasks. Identifying the brain networks that control particular WM subprocesses from a sea of parallel activity is a challenging task, requiring both clever experimental design and detailed joint brain-behavior modeling. It may also be the case that there is no clear correspondence between how a cognitive process is implemented at the neural level and its formal abstraction in a particular cognitive model. In such cases, we risk being blinded by the limited scope of our own models (i.e., there are no guarantees that the true model will be contained in the set of models we actually build). It is thus important to explore a large space of model variants across multiple levels of abstraction (Marr, 1982) (see Chapter "Linking Models with Brain Measures") in order to converge on the so-called ground truth of the neural systems of WM.

4 Concluding Remarks

In addition to the ongoing investigations in WM subprocesses outlined above, there are currently a number of "hot topics" in WM research worth mentioning that explore the relation between WM and other cognitive systems.

One such topic is *cognitive branching*, which refers to the ability to postpone execution of a primary task until a secondary task has been completed (Hyafil & Koechlin, 2016). Cognitive branching is crucial for goal-directed behavior because it enables flexible switching between different *task sets* (i.e., contextually appropriate configurations of cognitive processes used to perceive and respond to the environment) depending on the immediate context in which the person is acting (cf. Broadbent, 1970, 1971; Kahneman, 1973; for a review, see Sakai (2008)). This work has shown that only one cognitive branch (task set) can be active at once, meaning that our actions are governed by one plan at a time. Moreover, previously active branches are maintained in a "quiet" inactive state, which can be reactivated upon completion of the focal branch. Lateral prefrontal and frontopolar cortex have been implicated in switching between branches as goals and task demands change (Hyafil & Koechlin, 2016). EEG work shows that active and inactive WM items are encoded in dissociable neural patterns and that latent mnemonic representations can be reconfigured into an active action-ready state as task demands change (Muhle-Karbe et al., 2021). This ability to postpone contextually relevant information and action plans – and reactivate them at the appropriate time – also underlies *prospective memory*, which is the ability to maintain deferred intentions until the appropriate time or context occurs in the future (Boag et al., 2019a, b; Einstein & McDaniel, 2005; Hyafil & Koechlin, 2016; Strickland et al., 2018). Cognitive branching is

thought to be a domain general cognitive function central to multitasking and performing tasks that do not follow a pre-established plan. It has been suggested that selecting (and switching between) contextually appropriate branches or task sets is controlled by brain areas that track contextual cues such as uncertainty in expected reward and stimulus-response-outcome mappings (for a review, see Bland and Schaefer (2012)). In short, global changes in the external environment serve as an "alarm system" that signals when to switch to a more appropriate strategy (Bland & Schaefer, 2012). These ideas are consistent with broader "meta-control" theories of cognition that involve adopting contextually appropriate "processing styles" (e.g., Hommel (2020)) as well as Miller et al.'s (1960) original conception of WM as supporting the execution of multiple possible "action plans."

A second active area of research concerns the interplay between WM and reinforcement learning (RL) (i.e., learning driven by rewards and punishments, Sutton & Barto, 2018). Recent work has shown that participants rely on a mixture of "automatic" feedback-driven learning and effortful WM-based processing during instrumental learning (Collins & Frank, 2012; McDougle & Collins, 2021). Here, the WM and RL systems operate in concert during learning: participants use both short-term memory of recent outcomes and long-term item-value associations consolidated through error-driven learning to guide behavior (i.e., action selection). The relative weighting of each system in guiding behavior is modulated by WM load: when the number of items to learn is small (i.e., within capacity), the WM system dominates, whereas when the number of items to learn is large (i.e., capacity is exceeded), the RL system dominates (McDougle & Collins, 2021). This is consistent with the distinction between *model-based* and *model-free* learning in the RL literature (Doll et al., 2015; Otto et al., 2013), as well as broader *dual systems* theories of cognitive control that involve an interplay between automatic, associative processing (e.g., RL) and controlled, capacity-limited processing (e.g., WM) (Braver, 2012; Evans, 2003; Kahneman, 1973). Crucially, this work highlights the reciprocal relationship between WM and RL: WM supports learning by holding representations that guide value-based decision-making and are, in turn, updated by the learning system (Bakkour et al., 2018; Shadlen & Shohamy, 2016; Shohamy & Daw, 2015). WM subprocesses also "learn," for example, the timing of updating being trained by striatal "go/no-go" signaling (Frank et al., 2001; Hazy et al., 2006; O'Reilly & Frank, 2006).

A final interesting line of research concerns the interaction between WM, LTM, and perception. Verschooren et al. (2021) developed a modified reference-back paradigm in which items are gated into WM from either LTM or perception, which allows for comparison of gating dynamics for perceptual versus long-term memory information. This work found evidence that a single gate and a shared attentional selection mechanism control the access to WM for both perceptual and LTM information sources. Moreover, there are asymmetric costs associated with switching attention from internal (mnemonic) representations to external perceptual information and vice versa (Verschooren et al., 2020; for a review, see Verschooren et al. (2019); see also Roth and Courtney (2007)). In related work, Bartsch and Shepherdson (2021) showed that offloading items from WM to LTM reduces WM

load. Recall performance in a WM task was unaffected when memory load increased through the addition of items stored in LTM but deteriorated with the addition of items stored in WM, suggesting that individuals can flexibly draw on representations held in LTM in order to support WM-based decisions. Overall, the work discussed here highlights the importance of viewing WM not as an isolated module but as part of a broader set of complimentary cognitive systems (e.g., attention, perception, RL, LTM) that work together in order to support flexible goal-directed behavior. A promising avenue of future research would be to extend joint brain-behavior modeling approaches to account for data from multiple tasks, each of which may be explained by a different cognitive model or set of cognitive processes (e.g., Wall et al. (2021)).

In this chapter, we introduced the concept of WM and outlined key behavioral and neural evidence for a number of critical WM subprocesses. We highlighted several common approaches to linking brain and behavior in WM research and drew attention to limitations associated with common approaches. In doing so, we suggested that a number of recent advances in model-based cognitive neuroscience discussed throughout this book hold great promise in advancing our understanding of WM at the neural, cognitive, and behavioral levels. Overall, we are confident that such approaches will bring us closer to a unified model of WM that explains both behavioral and neural data in terms of a common set of neurocognitive mechanisms.

A.1 Exercises

1. A long-standing debate in the WM literature concerns whether items in WM are stored in a small number of discrete "slots" or allocated continuous resources from a limited capacity pool. Having read the review by Ma et al. (2014), how might you use neural data (e.g., EEG, fMRI) to distinguish between discrete slots and continuous resource models? What are two empirical findings that argue against the discrete slots perspective? Under what circumstances might discrete slots and continuous resource models mimic each other?
2. Concerning WM's severe capacity limit, Cowan (2010) says that two camps of theorists have emerged: one views capacity limits as a weakness, while the other views capacity limits as a strength. What are some of the arguments for capacity limits being a weakness? And for being a strength? Do you agree with this dichotomy? Can you think of any strengths or weaknesses that are not mentioned?
3. Cowan (2010) also mentions two phenomena that must be controlled for in WM experiments in order to obtain a pure measure of WM's capacity limit. What are the two phenomena and what are some methodological or experimental design strategies that are used to mitigate their confounding influence?
4. Another long-standing debate concerns whether mnemonic representations degrade over time due to temporal decay or interference from other sources.

Having read the review by Ricker et al. (2016), outline some evidence both for and against temporal decay. These authors state that "the biggest hurdle to providing evidence for or against the existence of decay has been preventing verbal rehearsal of memoranda while at the same time avoiding the introduction of interference." Why is this? How do these two confounds compare to those of Cowan (2010) identified in the previous question? (Hint: They are very similar!).

5. Considering what you have learned from this and other modeling chapters in this book, briefly sketch out a design for a cognitive model of WM that accounts for some of the subprocesses and behavioral phenomena discussed in this chapter. What kind of modeling architecture might you use as a starting point (e.g., neural network, evidence accumulation, etc.)? What mechanisms might be useful for explaining your target effects? How would you connect your model to neural data (e.g., EEG, fMRI)? Which neural signals might you link to which cognitive mechanisms and why?

B.1 Further Reading

- The classic review by Goldman-Rakic (1995) outlines evidence for the cellular basis of WM with a focus on early research using direct neuronal recordings in behaving animals.
- Cowan's classic (2001) paper provides detailed behavioral evidence for the "magical number 4" as WM's item capacity limit. Cowan (2010) provides additional interesting discussion and reflections on the capacity debate.
- Ma et al. (2014) review behavioral and neural evidence in support of the idea that WM relies on flexible allocation of continuous resources to keep items in an activated state.
- D'Esposito (2007) provides an interesting theoretical discussion of WM from a cognitive neuroscience perspective.
- Lewis-Peacock et al. (2018) discuss recent evidence for a number of distinct memory removal processes in the context of current cognitive and neural theories of WM.
- Ricker et al. (2016) give a historical overview of the debate surrounding whether mnemonic representations degrade over time due to temporal decay or interference from other items.

References

Anderson, M. L. (2014). *After phrenology: Neural reuse and the interactive brain*. MIT Press.
Anderson, M. C., & Hanslmayr, S. (2014). Neural mechanisms of motivated forgetting. *Trends in Cognitive Sciences, 18*(6), 279–292.
Anderson, J. R., & Lebiere, C. J. (2014). *The atomic components of thought*. Psychology Press.

Anderson, J. R., Reder, L. M., & Lebiere, C. (1996). Working memory: Activation limitations on retrieval. *Cognitive Psychology, 30*(3), 221–256.

Baddeley, A. (1992). Working memory. *Science, 255*(5044), 556–559. https://doi.org/10.1126/science.1736359

Badre, D. (2012). Opening the gate to working memory. *Proceedings of the National Academy of Sciences, 109*(49), 19878–19879.

Bakkour, A., Zylberberg, A., Shadlen, M. N., & Shohamy, D. (2018). Value-based decisions involve sequential sampling from memory. *BioRxiv*, 269290. https://doi.org/10.1101/269290

Bartsch, L. M., & Shepherdson, P. (2021). Freeing capacity in WM through the use of LTM representations. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 48*(4), 1–51.

Bays, P. M., & Husain, M. (2008). Dynamic shifts of limited working memory resources in human vision. *Science, 321*(5890), 851–854.

Bays, P. M., Catalao, R. F. G., & Husain, M. (2009). The precision of visual working memory is set by allocation of a shared resource. *Journal of Vision, 9*(10), 7.

Behrmann, M., Geng, J. J., & Shomstein, S. (2004). Parietal cortex and attention. *Current Opinion in Neurobiology, 14*(2), 212–217.

Bettencourt, K., & Xu, Y. (2015). Understanding the nature of visual short-term memory representation in human parietal cortex. *Journal of Vision, 15*(12), 292.

Bland, A. R., & Schaefer, A. (2012). Unexpected uncertainty, volatility and decision-making. *Frontiers in Neuroscience, 6*, 85.

Bledowski, C., Rahm, B., & Rowe, J. B. (2009). What "works" in working memory? Separate systems for selection and updating of critical information. *Journal of Neuroscience, 29*(43), 13735–13741.

Bledowski, C., Kaiser, J., & Rahm, B. (2010). Basic operations in working memory: Contributions from functional imaging studies. *Behavioural Brain Research, 214*, 172–179. https://doi.org/10.1016/j.bbr.2010.05.041

Boag, R. J., Strickland, L., Heathcote, A., Neal, A., & Loft, S. (2019a). Cognitive control and capacity for prospective memory in complex dynamic environments. *Journal of Experimental Psychology: General, 148*(12), 2181–2206. https://doi.org/10.1037/xge0000599

Boag, R. J., Strickland, L., Loft, S., & Heathcote, A. (2019b). Strategic attention and decision control support prospective memory in a complex dual-task environment. *Cognition, 191*, 103974. https://doi.org/10.1016/j.cognition.2019.05.011

Bogacz, R., Wagenmakers, E. J., Forstmann, B. U., & Nieuwenhuis, S. (2010). The neural basis of the speed-accuracy tradeoff. *Trends in Neurosciences, 33*, 10–16. https://doi.org/10.1016/j.tins.2009.09.002

Bor, D., Cumming, N., Scott, C. E. L., & Owen, A. M. (2004). Prefrontal cortical involvement in verbal encoding strategies. *European Journal of Neuroscience, 19*(12), 3365–3370.

Bouchacourt, F., Palminteri, S., Koechlin, E., & Ostojic, S. (2020). Temporal chunking as a mechanism for unsupervised learning of task-sets. *eLife, 9*, e50469.

Brady, T. F., Konkle, T., & Alvarez, G. A. (2009). Compression in visual working memory: Using statistical regularities to form more efficient memory representations. *Journal of Experimental Psychology: General, 138*(4), 487.

Braver, T. S. (2012). The variable nature of cognitive control: A dual mechanisms framework. *Trends in Cognitive Sciences, 16*, 105–112. https://doi.org/10.1016/j.tics.2011.12.010

Broadbent, D. E. (1970). Stimulus set and response set, two kinds of selective attention. In D. I. Mostofsky (Ed.), *Attention: Contemporary theory and analysis* (pp. 51–60). Appleton-Century-Crofts.

Broadbent, D. E. (1971). *Decision and stress*. Academic Press.

Brown, S. D., & Heathcote, A. (2008). The simplest complete model of choice response time: Linear ballistic accumulation. *Cognitive Psychology, 57*(3), 153–178. https://doi.org/10.1016/j.cogpsych.2007.12.002

Cantor, J., & Engle, R. W. (1993). Working-memory capacity as long-term memory activation: An individual-differences approach. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19*(5), 1101.

Cary, M., & Carlson, R. A. (2001). Distributing working memory resources during problem solving. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27*(3), 836.

Cavanagh, J. F. (2015). Cortical delta activity reflects reward prediction error and related behavioral adjustments, but at different times. *NeuroImage, 110*, 205–216.

Cavanagh, J. F., & Frank, M. J. (2014). Frontal theta as a mechanism for cognitive control. *Trends in Cognitive Sciences, 18*(8), 414–421.

Chatham, C. H., & Badre, D. (2015). Multiple gates on working memory. *Current Opinion in Behavioral Sciences, 1*, 23–31.

Clapp, W. C., Rubens, M. T., & Gazzaley, A. (2010). Mechanisms of working memory disruption by external interference. *Cerebral Cortex, 20*(4), 859–872.

Cohen, M. X. (2014). A neural microcircuit for cognitive conflict detection and signaling. *Trends in Neurosciences, 37*(9), 480–490.

Cohen, M. X., & Donner, T. H. (2013). Midfrontal conflict-related theta-band power reflects neural oscillations that predict behavior. *Journal of Neurophysiology, 110*(12), 2752–2763.

Collins, A. G. E., & Frank, M. J. (2012). How much of reinforcement learning is working memory, not reinforcement learning? A behavioral, computational, and neurogenetic analysis. *European Journal of Neuroscience, 35*(7), 1024–1035.

Cools, R. (2019). Chemistry of the adaptive mind: Lessons from dopamine. *Neuron, 104*, 113–131. https://doi.org/10.1016/j.neuron.2019.09.035

Cools, R., & D'Esposito, M. (2011). Inverted-U-shaped dopamine actions on human working memory and cognitive control. *Biological Psychiatry, 69*(12), e113–e125.

Cools, R., Lewis, S. J. G., Clark, L., Barker, R. A., & Robbins, T. W. (2007a). L-DOPA disrupts activity in the nucleus accumbens during reversal learning in Parkinson's disease. *Neuropsychopharmacology, 32*(1), 180–189. https://doi.org/10.1038/sj.npp.1301153

Cools, R., Sheridan, M., Jacobs, E., & D'Esposito, M. (2007b). Impulsive personality predicts dopamine-dependent changes in frontostriatal activity during component processes of working memory. *Journal of Neuroscience, 27*(20), 5506–5514.

Corrado, G., & Doya, K. (2007). Understanding neural coding through the model-based analysis of decision making. *The Journal of Neuroscience, 27*(31), 8178–8180. https://doi.org/10.1523/JNEUROSCI.1590-07.2007

Cowan, N. (1988). Evolving conceptions of memory storage, selective attention, and their mutual constraints within the human information-processing system. *Psychological Bulletin, 104*, 163–191.

Cowan, N. (1999). An embedded-processes model of working memory. In *Models of working memory: Mechanisms of active maintenance and executive control* (pp. 62–101). Cambridge University Press.

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences, 24*(1), 87–114. https://doi.org/10.1017/S0140525X01003922

Cowan, N. (2008). What are the differences between long-term, short-term, and working memory? *Progress in Brain Research, 169*, 323–338.

Cowan, N. (2010). The magical mystery four: How is working memory capacity limited, and why? *Current Directions in Psychological Science, 19*(1), 51–57.

Cowan, N. (2016). *Working memory capacity: Classic edition*. Psychology Press.

Cowan, N. (2017). The many faces of working memory and short-term storage. *Psychonomic Bulletin & Review, 24*(4), 1158–1170.

Cowan, N. (2019). Short-term memory based on activated long-term memory: A review in response to Norris (2017). *Psychological Bulletin, 145*(8), 822–847.

Cowan, N., Li, D., Moffitt, A., Becker, T. M., Martin, E. A., Saults, J. S., & Christ, S. E. (2011). A neural region of abstract working memory. *Journal of Cognitive Neuroscience, 23*(10), 2852–2863.

Curtis, C. E., & D'Esposito, M. (2003). Persistent activity in the prefrontal cortex during working memory. *Trends in Cognitive Sciences, 7*(9), 415–423.

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.

D'Esposito, M., & Postle, B. R. (2015). The cognitive neuroscience of working memory. *Annual Review of Psychology, 66*, 115–142.

D'Esposito, M., Postle, B. R., Ballard, D., & Lease, J. (1999). Maintenance versus manipulation of information held in working memory: An event-related fMRI study. *Brain and Cognition, 41*(1), 66–86.

Dahlin, E., Neely, A. S., Larsson, A., Bäckman, L., & Nyberg, L. (2008). Transfer of learning after updating training mediated by the striatum. *Science, 320*(5882), 1510–1512.

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior, 19*(4), 450–466. https://doi.org/10.1016/s0022-5371(80)90312-6

de Hollander, G., Keuken, M. C., van der Zwaag, W., Forstmann, B. U., & Trampel, R. (2017). Comparing functional MRI protocols for small, iron-rich basal ganglia nuclei such as the subthalamic nucleus at 7 T and 3 T. *Human Brain Mapping, 38*(6), 3226–3248. https://doi.org/10.1002/hbm.23586

Diamond, A. (2013). Executive functions. *Annual Review of Psychology, 64*, 135–168.

Ditterich, J. (2010). A comparison between mechanisms of multi-alternative perceptual decision making: Ability to explain human behavior, predictions for neurophysiology, and relationship with decision theory. *Frontiers in Neuroscience, 4*, 184.

Dolcos, F., Miller, B., Kragel, P., Jha, A., & McCarthy, G. (2007). Regional brain differences in the effect of distraction during the delay interval of a working memory task. *Brain Research, 1152*, 171–181.

Doll, B. B., Duncan, K. D., Simon, D. A., Shohamy, D., & Daw, N. D. (2015). Model-based choices involve prospective neural activity. *Nature Communications, 18*(5), 767–772. https://doi.org/10.1038/nn.3981

Donders, F. C. (1969). On the speed of mental processes. *Acta Psychologica, 30*(C), 412–431. https://doi.org/10.1016/0001-6918(69)90065-1

Donkin, C., Nosofsky, R. M., Gold, J. M., & Shiffrin, R. M. (2013). Discrete-slots models of visual working-memory response times. *Psychological Review, 120*(4), 873.

Donkin, C., Kary, A., Tahir, F., & Taylor, R. (2016). Resources masquerading as slots: Flexible allocation of visual working memory. *Cognitive Psychology, 85*, 30–42.

Dosenbach, N. U. F., Fair, D. A., Cohen, A. L., Schlaggar, B. L., & Petersen, S. E. (2008). A dual-networks architecture of top-down control. *Trends in Cognitive Sciences, 12*(3), 99–105.

Dreisbach, G. (2012). Mechanisms of cognitive control. *Current Directions in Psychological Science, 21*(4), 227–231. https://doi.org/10.1177/0963721412449830

Dreisbach, G., & Fröber, K. (2019). On how to be flexible (or not): Modulation of the stability-flexibility balance. *Current Directions in Psychological Science, 28*(1), 3–9. https://doi.org/10.1177/0963721418800030

Dreisbach, G., Müller, J., Goschke, T., Strobel, A., Schulze, K., Lesch, K.-P., & Brocke, B. (2005). Dopamine and cognitive control: The influence of spontaneous eyeblink rate and dopamine gene polymorphisms on perseveration and distractibility. *Behavioral Neuroscience, 119*(2), 483.

Durstewitz, D., & Seamans, J. K. (2002). The computational role of dopamine D1 receptors in working memory. *Neural Networks, 15*(4–6), 561–572. https://doi.org/10.1016/S0893-6080(02)00049-7

Durstewitz, D., & Seamans, J. K. (2008). The dual-state theory of prefrontal cortex dopamine function with relevance to catechol-O-methyltransferase genotypes and schizophrenia. *Biological Psychiatry, 64*, 739–749. https://doi.org/10.1016/j.biopsych.2008.05.015

Durstewitz, D., Seamans, J. K., & Sejnowski, T. J. (2000). Dopamine-mediated stabilization of delay-period activity in a network model of prefrontal cortex. *Journal of Neurophysiology, 83*(3), 1733–1750.

Ecker, U. K. H., Lewandowsky, S., Oberauer, K., & Chee, A. E. H. (2010). The components of working memory updating: An experimental decomposition and individual differences. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36*(1), 170.

Ecker, U. K. H., Lewandowsky, S., & Oberauer, K. (2014a). 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

Ecker, U. K. H., Oberauer, K., & Lewandowsky, S. (2014b). Working memory updating involves item-specific removal. *Journal of Memory and Language, 74*, 1–15. https://doi.org/10.1016/j.jml.2014.03.006

Einstein, G. O., & McDaniel, M. A. (2005). Prospective memory: Multiple retrieval processes. *Current Directions in Psychological Science, 14*(6), 286–290.

Engle, R. W., Tuholski, S. W., Laughlin, J. E., & Conway, A. R. A. (1999). Working memory, short-term memory, and general fluid intelligence: A latent-variable approach. *Journal of Experimental Psychology: General, 128*(3), 309.

Evans, J. S. B. T. (2003). In two minds: Dual-process accounts of reasoning. *Trends in Cognitive Sciences, 7*(10), 454–459.

Farrell, S., & Lewandowsky, S. (2002). An endogenous distributed model of ordering in serial recall. *Psychonomic Bulletin & Review, 9*(1), 59–79.

Feredoes, E., Heinen, K., Weiskopf, N., Ruff, C., & Driver, J. (2011). Causal evidence for frontal involvement in memory target maintenance by posterior brain areas during distracter interference of visual working memory. *Proceedings of the National Academy of Sciences, 108*(42), 17510–17515.

Festini, S. B., & Reuter-Lorenz, P. A. (2013). The short-and long-term consequences of directed forgetting in a working memory task. *Memory, 21*(7), 763–777.

Festini, S. B., & Reuter-Lorenz, P. A. (2014). Cognitive control of familiarity: Directed forgetting reduces proactive interference in working memory. *Cognitive, Affective, & Behavioral Neuroscience, 14*(1), 78–89.

Folstein, J. R., & Van Petten, C. (2008). Influence of cognitive control and mismatch on the N2 component of the ERP: A review. *Psychophysiology, 45*(1), 152–170.

Forstmann, B. U., & Wagenmakers, E.-J. (2015). Model-based cognitive neuroscience: A conceptual introduction. In *An introduction to model-based cognitive neuroscience* (pp. 139–156). Springer. https://doi.org/10.1007/978-1-4939-2236-9_7

Forstmann, B. U., Dutilh, G., Brown, S., Neumann, J., Von Cramon, D. Y., Ridderinkhof, K. R., & Wagenmakers, E.-J. (2008). Striatum and pre-SMA facilitate decision-making under time pressure. *Proceedings of the National Academy of Sciences of the United States of America, 105*(45), 17538–17542. https://doi.org/10.1073/pnas.0805903105

Forstmann, B. U., Wagenmakers, E.-J., Eichele, T., Brown, S., & Serences, J. T. (2011). Reciprocal relations between cognitive neuroscience and formal cognitive models: Opposites attract? *Trends in Cognitive Sciences, 15*, 272–279. https://doi.org/10.1016/j.tics.2011.04.002

Forstmann, B. U., Ratcliff, R., & Wagenmakers, E.-J. (2016). Sequential sampling models in cognitive neuroscience: Advantages, applications, and extensions. *Annual Review of Psychology, 67*(1), 641–666. https://doi.org/10.1146/annurev-psych-122414-033645

Frank, M. J., Loughry, B., & O'Reilly, R. C. (2001). Interactions between frontal cortex and basal ganglia in working memory: A computational model. *Cognitive, Affective, & Behavioral Neuroscience, 1*, 137–160. https://doi.org/10.3758/CABN.1.2.137

Friston, K. J. (2009). Modalities, modes, and models in functional neuroimaging. *Science, 326*, 399–403. https://doi.org/10.1126/science.1174521

Funahashi, S., Bruce, C. J., & Goldman-Rakic, P. S. (1989). Mnemonic coding of visual space in the monkey's dorsolateral prefrontal cortex. *Journal of Neurophysiology, 61*(2), 331–349.

Funahashi, S., Bruce, C. J., & Goldman-Rakic, P. S. (1990). Visuospatial coding in primate prefrontal neurons revealed by oculomotor paradigms. *Journal of Neurophysiology, 63*(4), 814–831.

Funahashi, S., Bruce, C. J., & Goldman-Rakic, P. S. (1991). Neuronal activity related to saccadic eye movements in the monkey's dorsolateral prefrontal cortex. *Journal of Neurophysiology, 65*(6), 1464–1483.

Garavan, H. (1998). Serial attention within working memory. *Memory & Cognition, 26*(2), 263–276.

Gehring, W. J., Bryck, R. L., Jonides, J., Albin, R. L., & Badre, D. (2003). The mind's eye, looking inward? In search of executive control in internal attention shifting. *Psychophysiology, 40*(4), 572–585.

Goldman-Rakic, P. S. (1995). Cellular basis of working memory. *Neuron, 14*(3), 477–485.

Grech, R., Cassar, T., Muscat, J., Camilleri, K. P., Fabri, S. G., Zervakis, M., et al. (2008). Review on solving the inverse problem in EEG source analysis. *Journal of Neuroengineering and Rehabilitation, 5*(1), 1–33.

Gulbinaite, R., van Rijn, H., & Cohen, M. X. (2014). Fronto-parietal network oscillations reveal relationship between working memory capacity and cognitive control. *Frontiers in Human Neuroscience, 8*, 761.

Hardman, K. O., & Cowan, N. (2015). Remembering complex objects in visual working memory: Do capacity limits restrict objects or features? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 41*(2), 325.

Harmony, T. (2013). The functional significance of delta oscillations in cognitive processing. *Frontiers in Integrative Neuroscience, 7*, 83.

Hauswald, A., Schulz, H., Iordanov, T., & Kissler, J. (2011). ERP dynamics underlying successful directed forgetting of neutral but not negative pictures. *Social Cognitive and Affective Neuroscience, 6*(4), 450–459.

Hawkins, G. E., Mittner, M., Forstmann, B. U., & Heathcote, A. (2017). On the efficiency of neurally-informed cognitive models to identify latent cognitive states. *Journal of Mathematical Psychology, 76*, 142–155. https://doi.org/10.1016/j.jmp.2016.06.007

Hazy, T. E., Frank, M. J., & O'Reilly, R. C. (2006). Banishing the homunculus: Making working memory work. *Neuroscience, 139*(1), 105–118. https://doi.org/10.1016/j.neuroscience.2005.04.067

Hedge, C., & Leonards, U. (2013). Using eye movements to explore switch costs in working memory. *Journal of Vision, 13*(4), 18.

Hommel, B. (2015). Between persistence and flexibility: The Yin and Yang of action control. *Advances in Motivation Science, 2*, 33–67. https://doi.org/10.1016/bs.adms.2015.04.003

Hommel, B. (2020). Dual-task performance: Theoretical analysis and an event-coding account. *Journal of Cognition, 3*(1), 29.

Howard, M. W., Rizzuto, D. S., Caplan, J. B., Madsen, J. R., Lisman, J., Aschenbrenner-Scheibe, R., et al. (2003). Gamma oscillations correlate with working memory load in humans. *Cerebral Cortex, 13*(12), 1369–1374.

Hyafil, A., & Koechlin, E. (2016). A neurocomputational model of human frontopolar cortex function. *BioRxiv*, 37150. https://doi.org/10.1101/037150

Jocham, G., Klein, T. A., & Ullsperger, M. (2011). Dopamine-mediated reinforcement learning signals in the striatum and ventromedial prefrontal cortex underlie value-based choices. *Journal of Neuroscience, 31*(5), 1606–1613.

Jongkees, B. J. (2020). Baseline-dependent effect of dopamine's precursor L-tyrosine on working memory gating but not updating. *Cognitive, Affective, & Behavioral Neuroscience, 20*(3), 521–535. https://doi.org/10.3758/s13415-020-00783-8

Kahneman, D. (1973). *Attention and effort* (Vol. 1063). Citeseer.

Kessler, Y., & Meiran, N. (2008). Two dissociable updating processes in working memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34*(6), 1339.

Kessler, Y., & Oberauer, K. (2014). Working memory updating latency reflects the cost of switching between maintenance and updating modes of operation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 40*(3), 738.

Kessler, Y., & Oberauer, K. (2015). Forward scanning in verbal working memory updating. *Psychonomic Bulletin & Review, 22*(6), 1770–1776.

Kornblith, S., Buschman, T. J., & Miller, E. K. (2016). Stimulus load and oscillatory activity in higher cortex. *Cerebral Cortex, 26*(9), 3772–3784.

Kübler, A., Murphy, K., Kaufman, J., Stein, E. A., & Garavan, H. (2003). Co-ordination within and between verbal and visuospatial working memory: Network modulation and anterior frontal recruitment. *NeuroImage, 20*(2), 1298–1308.

Kyllonen, P. C., & Christal, R. E. (1990). Reasoning ability is (little more than) working-memory capacity?! *Intelligence, 14*(4), 389–433. https://doi.org/10.1016/S0160-2896(05)80012-1

Lepsien, J., Griffin, I. C., Devlin, J. T., & Nobre, A. C. (2005). Directing spatial attention in mental representations: Interactions between attentional orienting and working-memory load. *NeuroImage, 26*(3), 733–743.

Leung, H.-C., Seelig, D., & Gore, J. C. (2004). The effect of memory load on cortical activity in the spatial working memory circuit. *Cognitive, Affective, & Behavioral Neuroscience, 4*(4), 553–563.

Lewis, S. J. G., Dove, A., Robbins, T. W., Barker, R. A., & Owen, A. M. (2004). Striatal contributions to working memory: A functional magnetic resonance imaging study in humans. *European Journal of Neuroscience, 19*(3), 755–760.

Lewis-Peacock, J. A., Drysdale, A. T., Oberauer, K., & Postle, B. R. (2012). Neural evidence for a distinction between short-term memory and the focus of attention. *Journal of Cognitive Neuroscience, 24*(1), 61–79.

Lewis-Peacock, J. A., Kessler, Y., & Oberauer, K. (2018). The removal of information from working memory. *Annals of the New York Academy of Sciences, 1424*(1), 33–44.

Logie, R. H. (1995). *Visuo-spatial working memory*. Lawrence Erlbaum Associates, Inc.

Logie, R. H., Gilhooly, K. J., & Wynn, V. (1994). Counting on working memory in arithmetic problem solving. *Memory & Cognition, 22*(4), 395–410.

Lorenc, E. S., Sreenivasan, K. K., Nee, D. E., Vandenbroucke, A. R. E., & D'Esposito, M. (2018). Flexible coding of visual working memory representations during distraction. *Journal of Neuroscience, 38*(23), 5267–5276.

Love, B. C. (2016). Cognitive models as bridge between brain and behavior. *Trends in Cognitive Sciences, 20*, 247–248. https://doi.org/10.1016/j.tics.2016.02.006

Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature, 390*(6657), 279–281.

Luck, S. J., & Vogel, E. K. (2013). Visual working memory capacity: From psychophysics and neurobiology to individual differences. *Trends in Cognitive Sciences, 17*(8), 391–400.

Ma, W. J., Husain, M., & Bays, P. M. (2014). Changing concepts of working memory. *Nature Neuroscience, 17*(3), 347.

Mack, M. L., Preston, A. R., & Love, B. C. (2013). Decoding the brain's algorithm for categorization from its neural implementation. *Current Biology, 23*(20), 2023–2027.

Manoach, D. S., Schlaug, G., Siewert, B., Darby, D. G., Bly, B. M., Benfield, A., et al. (1997). Prefrontal cortex fMRI signal changes are correlated with working memory load. *Neuroreport, 8*(2), 545–549.

Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information*. W.H. Freeman.

Marshuetz, C. (2005). Order information in working memory: An integrative review of evidence from brain and behavior. *Psychological Bulletin, 131*(3), 323.

Mathy, F., & Feldman, J. (2012). What's magic about magic numbers? Chunking and data compression in short-term memory. *Cognition, 122*(3), 346–362.

McDougle, S. D., & Collins, A. G. E. (2021). Modeling the influence of working memory, reinforcement, and action uncertainty on reaction time and choice during instrumental learning. *Psychonomic Bulletin & Review, 28*, 20–39.

McNab, F., & Klingberg, T. (2008). Prefrontal cortex and basal ganglia control access to working memory. *Nature Neuroscience, 11*(1), 103–107.

Miletić, S., Bazin, P.-L., Weiskopf, N., van der Zwaag, W., Forstmann, B. U., & Trampel, R. (2020). fMRI protocol optimization for simultaneously studying small subcortical and cortical areas at 7 T. *NeuroImage, 219*, 116992.

Miller, E. K., & Cohen, J. D. (2001). An integrative theory of prefrontal cortex function. *Annual Review of Neuroscience, 24*(1), 167–202.
Miller, G. A., Galanter, E., & Pribram, K. H. (1960). *Plans and the structure of behavior*. Henry Holt and Co. https://doi.org/10.1037/10039-000
Miller, E. K., Lundqvist, M., & Bastos, A. M. (2018). Working Memory 2.0. *Neuron, 100*, 463–475. https://doi.org/10.1016/j.neuron.2018.09.023
Möller, M., & Bogacz, R. (2019). Learning the payoffs and costs of actions. *PLoS Computational Biology, 15*(2), e1006285. https://doi.org/10.1371/journal.pcbi.1006285
Morey, C. C., & Cowan, N. (2018). Can we distinguish three maintenance processes in working memory? *Annals of the New York Academy of Sciences, 1424*(1), 45.
Muhle-Karbe, P. S., Myers, N. E., & Stokes, M. G. (2021). A hierarchy of functional states in working memory. *Journal of Neuroscience, 41*(20), 4461–4475.
Murty, V. P., Sambataro, F., Radulescu, E., Altamura, M., Iudicello, J., Zoltick, B., et al. (2011). Selective updating of working memory content modulates meso-cortico-striatal activity. *NeuroImage, 57*(3), 1264–1272. https://doi.org/10.1016/j.neuroimage.2011.05.006
Musslick, S., Jang, S. J., Shvartsman, M., Shenhav, A., & Cohen, J. D. (2018). Constraints associated with cognitive control and the stability-flexibility dilemma. In *Proceedings of the 40th annual meeting of the Cognitive Science Society, CogSci 2018* (pp. 804–809). The Cognitive Science Society.
Nassar, M. R., Helmers, J. C., & Frank, M. J. (2018). Chunking as a rational strategy for lossy data compression in visual working memory. *Psychological Review, 125*(4), 486.
Nir-Cohen, G., Kessler, Y., & Egner, T. (2020). Neural substrates of working memory updating. *Journal of Cognitive Neuroscience, 32*(12), 2285–2302. https://doi.org/10.1162/jocn_a_01625
Nowicka, A., Marchewka, A., Jednorog, K., Tacikowski, P., & Brechmann, A. (2011). Forgetting of emotional information is hard: An fMRI study of directed forgetting. *Cerebral Cortex, 21*(3), 539–549.
O'Reilly, R. C. (2006). Biologically based computational models of high-level cognition. *Science, 314*, 91–94. https://doi.org/10.1126/science.1127242
O'Reilly, R. C., & Frank, M. J. (2006). Making working memory work: A computational model of learning in the prefrontal cortex and basal ganglia. *Neural Computation, 18*(2), 283–328. https://doi.org/10.1162/089976606775093909
O'Reilly, R. C., & Munakata, Y. (2000). *Computational explorations in cognitive neuroscience: Understanding the mind by simulating the brain*. MIT Press.
Oberauer, K. (2002). Access to information in working memory: Exploring the focus of attention. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28*(3), 411.
Oberauer, K. (2003). Selective attention to elements in working memory. *Experimental Psychology, 50*(4), 257.
Oberauer, K. (2006). Is the focus of attention in working memory expanded through practice? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32*(2), 197.
Oberauer, K. (2009). Design for a working memory. *Psychology of Learning and Motivation – Advances in Research and Theory, 51*, 45–100. https://doi.org/10.1016/S0079-7421(09)51002-X
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.
Oberauer, K. (2019). Working memory and attention–A conceptual analysis and review. *Journal of Cognition, 2*(1), 36.
Oberauer, K., & Bialkova, S. (2011). Serial and parallel processes in working memory after practice. *Journal of Experimental Psychology: Human Perception and Performance, 37*(2), 606.
Oberauer, K., & Lewandowsky, S. (2011). Modeling working memory: A computational implementation of the Time-Based Resource-Sharing theory. *Psychonomic Bulletin & Review, 18*(1), 10–45.

Oberauer, K., & Lewandowsky, S. (2016). Control of information in working memory: Encoding and removal of distractors in the complex-span paradigm. *Cognition, 156*, 106–128.

Oberauer, K., Lewandowsky, S., Awh, E., Brown, G. D. A., Conway, A., Cowan, N., et al. (2018). Benchmarks for models of short-term and working memory. *Psychological Bulletin, 144*(9), 885.

Ortega, R., López, V., Carrasco, X., Escobar, M. J., García, A. M., Parra, M. A., & Aboitiz, F. (2020). Neurocognitive mechanisms underlying working memory encoding and retrieval in Attention-Deficit/Hyperactivity Disorder. *Scientific Reports, 10*(1), 1–13.

Otto, A. R., Gershman, S. J., Markman, A. B., & Daw, N. D. (2013). The curse of planning: Dissecting multiple reinforcement-learning systems by taxing the central executive. *Psychological Science, 24*(5), 751–761.

Owen, A. M., McMillan, K. M., Laird, A. R., & Bullmore, E. (2005). N-back working memory paradigm: A meta-analysis of normative functional neuroimaging studies. *Human Brain Mapping, 25*(1), 46–59.

Pang, J., Tang, X., Nie, Q.-Y., Conci, M., Sun, P., Wang, H., et al. (2019). Resolving the electroencephalographic correlates of rapid goal-directed chunking in the frontal-parietal network. *Frontiers in Neuroscience, 13*, 744.

Parent, A., & Hazrati, L.-N. (1995). Functional anatomy of the basal ganglia. I. The cortico-basal ganglia-thalamo-cortical loop. *Brain Research Reviews, 20*(1), 91–127.

Paz-Caballero, M. D., Menor, J., & Jiménez, J. M. (2004). Predictive validity of event-related potentials (ERPs) in relation to the directed forgetting effects. *Clinical Neurophysiology, 115*(2), 369–377.

Pearson, B., Raškevičius, J., Bays, P. M., Pertzov, Y., & Husain, M. (2014). Working memory retrieval as a decision process. *Journal of Vision, 14*(2), 1–15. https://doi.org/10.1167/14.2.2

Purcell, B. A., & Palmeri, T. J. (2017). Relating accumulator model parameters and neural dynamics. *Journal of Mathematical Psychology, 76*, 156–171.

Purcell, B. A., Heitz, R. P., Cohen, J. Y., Schall, J. D., Logan, G. D., & Palmeri, T. J. (2010). Neurally constrained modeling of perceptual decision making. *Psychological Review, 117*(4), 1113–1143. https://doi.org/10.1037/a0020311

Rac-Lubashevsky, R., & Kessler, Y. (2016a). Decomposing the n-back task: An individual differences study using the reference-back paradigm. *Neuropsychologia, 90*, 190–199. https://doi.org/10.1016/j.neuropsychologia.2016.07.013

Rac-Lubashevsky, R., & Kessler, Y. (2016b). Dissociating working memory updating and automatic updating: The reference-back paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 42*(6), 951–969. https://doi.org/10.1037/xlm0000219

Rac-Lubashevsky, R., & Kessler, Y. (2018). Oscillatory correlates of control over working memory gating and updating: An EEG study using the reference-back paradigm. *Journal of Cognitive Neuroscience, 30*(12), 1870–1882. https://doi.org/10.1162/jocn_a_01326

Rac-Lubashevsky, R., & Kessler, Y. (2019). Revisiting the relationship between the P3b and working memory updating. *Biological Psychology, 148*, 107769. https://doi.org/10.1016/j.biopsycho.2019.107769

Rac-Lubashevsky, R., Slagter, H. A., & Kessler, Y. (2017). Tracking real-time changes in working memory updating and gating with the event-based eye-blink rate. *Scientific Reports, 7*(1), 1–9. https://doi.org/10.1038/s41598-017-02942-3

Rademaker, R. L., Tredway, C. H., & Tong, F. (2012). Introspective judgments predict the precision and likelihood of successful maintenance of visual working memory. *Journal of Vision, 12*(13), 21.

Ranganath, C., Cohen, M. X., & Brozinsky, C. J. (2005). Working memory maintenance contributes to long-term memory formation: Neural and behavioral evidence. *Journal of Cognitive Neuroscience, 17*(7), 994–1010.

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review, 8*, 59–108. Retrieved from https://psycnet.apa.org/record/1978-30970-001

Reber, P. J., Siwiec, R. M., Gitleman, D. R., Parrish, T. B., Mesulam, M.-M., & Paller, K. A. (2002). Neural correlates of successful encoding identified using functional magnetic resonance imaging. *Journal of Neuroscience, 22*(21), 9541–9548.

Ricker, T. J., Vergauwe, E., & Cowan, N. (2016). Decay theory of immediate memory: From Brown (1958) to today (2014). *Quarterly Journal of Experimental Psychology, 69*(10), 1969–1995.

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*(38), 12990–12998.

Rizio, A. A., & Dennis, N. A. (2013). The neural correlates of cognitive control: Successful remembering and intentional forgetting. *Journal of Cognitive Neuroscience, 25*(2), 297–312.

Roth, J. K., & Courtney, S. M. (2007). Neural system for updating object working memory from different sources: Sensory stimuli or long-term memory. *NeuroImage, 38*(3), 617–630.

Roth, J. K., Serences, J. T., & Courtney, S. M. (2006). Neural system for controlling the contents of object working memory in humans. *Cerebral Cortex, 16*, 1595–1603. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.524.778

Roux, F., & Uhlhaas, P. J. (2014). Working memory and neural oscillations: Alpha–gamma versus theta–gamma codes for distinct WM information? *Trends in Cognitive Sciences, 18*(1), 16–25.

Roux, F., Wibral, M., Mohr, H. M., Singer, W., & Uhlhaas, P. J. (2012). Gamma-band activity in human prefrontal cortex codes for the number of relevant items maintained in working memory. *Journal of Neuroscience, 32*(36), 12411–12420.

Sakai, K. (2008). Task set and prefrontal cortex. *Annual Review of Neuroscience, 31*, 219–245.

Sakai, K., Rowe, J. B., & Passingham, R. E. (2002). Active maintenance in prefrontal area 46 creates distractor-resistant memory. *Nature Neuroscience, 5*(5), 479–484.

Santangelo, V., Di Francesco, S. A., Mastroberardino, S., & Macaluso, E. (2015). Parietal cortex integrates contextual and saliency signals during the encoding of natural scenes in working memory. *Human Brain Mapping, 36*(12), 5003–5017.

Schall, J. D. (2019). Accumulators, neurons, and response time. *Trends in Neurosciences, 42*, 848–860. https://doi.org/10.1016/j.tins.2019.10.001

Schultz, W. (1998). Predictive reward signal of dopamine neurons. *Journal of Neurophysiology, 80*, 1–27. https://doi.org/10.1152/jn.1998.80.1.1

Schultz, W., Dayan, P., & Montague, P. R. (1997). A neural substrate of prediction and reward. *Science, 275*(5306), 1593–1599. https://doi.org/10.1126/science.275.5306.1593

Sewell, D. K., Lilburn, S. D., & Smith, P. L. (2014). An information capacity limitation of visual short-term memory. *Journal of Experimental Psychology: Human Perception and Performance, 40*(6), 2214.

Sewell, D. K., Lilburn, S. D., & Smith, P. L. (2016). Object selection costs in visual working memory: A diffusion model analysis of the focus of attention. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 42*(11), 1673.

Shadlen, M. N., & Shohamy, D. (2016). Decision making and sequential sampling from memory. *Neuron, 90*(5), 927–939.

Shiffrin, R. M., & Atkinson, R. C. (1969). Storage and retrieval processes in long-term memory. *Psychological Review, 76*(2), 179.

Shohamy, D., & Daw, N. D. (2015). Integrating memories to guide decisions. *Current Opinion in Behavioral Sciences, 5*, 85–90.

Silver, H., Feldman, P., Bilker, W., & Gur, R. C. (2003). Working memory deficit as a core neuropsychological dysfunction in schizophrenia. *American Journal of Psychiatry, 160*(10), 1809–1816.

Strickland, L., Loft, S., Remington, R. W., & Heathcote, A. (2018). Racing to remember: A theory of decision control in event-based prospective memory. *Psychological Review, 125*(6), 851–887. https://doi.org/10.1037/rev0000113

Sutton, R. S., & Barto, A. G. (2018). *Reinforcement learning: An introduction* (2nd ed.). MIT Press.

Tan, H.-Y., Chen, Q., Goldberg, T. E., Mattay, V. S., Meyer-Lindenberg, A., Weinberger, D. R., & Callicott, J. H. (2007). Catechol-O-methyltransferase Val158Met modulation of prefrontal–parietal–striatal brain systems during arithmetic and temporal transformations in working memory. *Journal of Neuroscience, 27*(49), 13393–13401.

Thalmann, M., Souza, A. S., & Oberauer, K. (2019). How does chunking help working memory? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 45*(1), 37.

Timm, J. D., & Papenmeier, F. (2019). Reorganization of spatial configurations in visual working memory. *Memory & Cognition, 47*(8), 1469–1480.

Todd, J. J., & Marois, R. (2004). Capacity limit of visual short-term memory in human posterior parietal cortex. *Nature, 428*(6984), 751–754.

Toepper, M., Gebhardt, H., Beblo, T., Thomas, C., Driessen, M., Bischoff, M., et al. (2010). Functional correlates of distractor suppression during spatial working memory encoding. *Neuroscience, 165*(4), 1244–1253.

Trutti, A. C., Mulder, M. J., Hommel, B., & Forstmann, B. U. (2019). Functional neuroanatomical review of the ventral tegmental area. *NeuroImage, 191*, 258–268.

Trutti, A. C., Verschooren, S., Forstmann, B. U., & Boag, R. J. (2021). Understanding subprocesses of working memory through the lens of model-based cognitive neuroscience. *Current Opinion in Behavioral Sciences, 38*, 57–65. https://doi.org/10.1016/j.cobeha.2020.10.002

Turner, B. M., Forstmann, B. U., Wagenmakers, E. J., Brown, S. D., Sederberg, P. B., & Steyvers, M. (2013). A Bayesian framework for simultaneously modeling neural and behavioral data. *NeuroImage, 72*, 193–206. https://doi.org/10.1016/j.neuroimage.2013.01.048

Turner, B. M., Rodriguez, C. A., Norcia, T. M., McClure, S. M., & Steyvers, M. (2016). Why more is better: Simultaneous modeling of EEG, fMRI, and behavioral data. *NeuroImage, 128*, 96–115.

Turner, B. M., Forstmann, B. U., Love, B. C., Palmeri, T. J., & Van Maanen, L. (2017a). Approaches to analysis in model-based cognitive neuroscience. *Journal of Mathematical Psychology, 76*, 65–79. https://doi.org/10.1016/j.jmp.2016.01.001

Turner, B. M., Wang, T., & Merkle, E. C. (2017b). Factor analysis linking functions for simultaneously modeling neural and behavioral data. *NeuroImage, 153*, 28–48. https://doi.org/10.1016/j.neuroimage.2017.03.044

Turner, B. M., Forstmann, B. U., & Steyvers, M. (2019a). A tutorial on joint modeling. In *Joint models of neural and behavioral data. Computational approaches to cognition and perception*. Springer. https://doi.org/10.1007/978-3-030-03688-1_2

Turner, B. M., Palestro, J. J., Miletić, S., & Forstmann, B. U. (2019b). Advances in techniques for imposing reciprocity in brain-behavior relations. *Neuroscience and Biobehavioral Reviews, 102*, 327–336. https://doi.org/10.1016/j.neubiorev.2019.04.018

van Dijck, J.-P., Abrahamse, E. L., Majerus, S., & Fias, W. (2013). Spatial attention interacts with serial-order retrieval from verbal working memory. *Psychological Science, 24*(9), 1854–1859.

van Schouwenburg, M. R., den Ouden, H. E. M., & Cools, R. (2010). The human basal ganglia modulate frontal-posterior connectivity during attention shifting. *Journal of Neuroscience, 30*(29), 9910–9918.

Veltman, D. J., Rombouts, S. A. R. B., & Dolan, R. J. (2003). Maintenance versus manipulation in verbal working memory revisited: An fMRI study. *NeuroImage, 18*(2), 247–256.

Verschooren, S., Schindler, S., De Raedt, R., & Pourtois, G. (2019). Switching attention from internal to external information processing: A review of the literature and empirical support of the resource sharing account. *Psychonomic Bulletin & Review, 26*(2), 468–490.

Verschooren, S., Pourtois, G., & Egner, T. (2020). More efficient shielding for internal than external attention? Evidence from asymmetrical switch costs. *Journal of Experimental Psychology: Human Perception and Performance, 46*(9), 912–925.

Verschooren, S., Kessler, Y., & Egner, T. (2021). Evidence for a single mechanism gating perceptual and long-term memory information into working memory. *Cognition, 212*, 104668. https://doi.org/10.1016/j.cognition.2021.104668

Vogel, E. K., & Machizawa, M. G. (2004). Neural activity predicts individual differences in visual working memory capacity. *Nature, 428*(6984), 748–751.

Wagenmakers, E.-J., Van Der Maas, H. L. J., & Grasman, R. P. P. P. (2007). An EZ-diffusion model for response time and accuracy. *Psychonomic Bulletin & Review, 14*(1), 3–22.

Wall, L., Gunawan, D., Brown, S. D., Tran, M.-N., Kohn, R., & Hawkins, G. E. (2021). Identifying relationships between cognitive processes across tasks, contexts, and time. *Behavior Research Methods, 53*(1), 78–95.

Wilson, M. (2002). Six views of embodied cognition. *Psychonomic Bulletin & Review, 9*(4), 625–636.

Wylie, G. R., Foxe, J. J., & Taylor, T. L. (2008). Forgetting as an active process: An fMRI investigation of item-method–directed forgetting. *Cerebral Cortex, 18*(3), 670–682.

Xu, Y., & Chun, M. M. (2006). Dissociable neural mechanisms supporting visual short-term memory for objects. *Nature, 440*(7080), 91–95.

Zhang, W., & Luck, S. J. (2008). Discrete fixed-resolution representations in visual working memory. *Nature, 453*(7192), 233–235.

Zhao, Y., Kuai, S., Zanto, T. P., & Ku, Y. (2020). Neural correlates underlying the precision of visual working memory. *Neuroscience, 425*, 301–311.