



Universiteit
Leiden

The Netherlands

Investigating structure and function of the dopaminergic midbrain: with a special focus on the human VTA

Trutti, A.C.

Citation

Trutti, A. C. (2024, November 15). *Investigating structure and function of the dopaminergic midbrain: with a special focus on the human VTA*. Retrieved from <https://hdl.handle.net/1887/4108996>

Version: Publisher's Version

License: [Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden](#)

Downloaded from: <https://hdl.handle.net/1887/4108996>

Note: To cite this publication please use the final published version (if applicable).

Chapter 5

Understanding subprocesses of working memory through the lens of model-based cognitive neuroscience

This chapter is published as:

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 Behavioural Sciences*, 38, 57–65. <https://doi.org/10.1016/j.cobeha.2020.10.002>

Abstract

Working memory (working memory) refers to a set of processes that makes task-relevant information accessible to higher-level cognitive processes. Recent work suggests working memory is supported by a variety of information gating, updating, and removal processes, which ensure only task-relevant information occupies working memory. Current neurocomputational theory suggests working memory gating is accomplished via 'go/no-go' signalling in basal ganglia-thalamus-prefrontal cortex pathways but is less clear about other subprocesses and brain structures known to play a role in working memory. We review recent efforts to identify the neural basis of working memory subprocesses using the recently developed reference-back task as a benchmark measure of working memory subprocesses. Targets for future research using the methods of model-based cognitive neuroscience and novel extensions to the reference-back task are suggested.

5.1 Working memory and its subprocesses

Working memory refers to a set of processes that makes task-relevant information accessible to higher-level cognitive processes such as learning, decision-making, reasoning, and reading comprehension (Cowan, 1988; Daneman & Carpenter, 1980; Kyllonen & Christal, 1990). Working memory is extremely capacity-limited, with current research suggesting that between one and four items¹ can be maintained in an activated state in working memory at a time (Cowan, 2001; Garavan, 1998; Oberauer, 2002; Sewell et al., 2016). This strict limit demands a high degree of control over working memory content, such that working memory must strike a balance between stability (i.e., protecting the current contents of working memory from irrelevant or distracting information) and flexibility (i.e., keeping working memory up to date with new relevant information and removing outdated information). This trade-off between stability and flexibility (Dreisbach, 2012; Dreisbach & Fröber, 2019; Dreisbach et al., 2005; Hommel, 2015) is a core feature of executive control processes (e.g., cognitive control, conflict monitoring/resolution, task switching; (Musslick et al., 2018) and managing the trade-off strongly depends on the brain's dopamine systems (Cools, 2019; Cools & D'Esposito, 2011).

Prominent computational theories suggest that working memory resolves the stability-flexibility trade-off by operating in two modes: An updating (gate-open) mode, which allows new information to enter working memory, and a maintenance (gate-closed) mode, which prevents irrelevant and distracting information from interfering with the current contents of working memory (Badre, 2012; Bledowski et al., 2010; Kessler & Oberauer, 2014; Miller & Cohen, 2001; Murty et al., 2011; O'Reilly, 2006; Rac-Lubashevsky & Kessler, 2016b; Roth et al., 2006). In the gate-open mode, updating is further supported by two main subprocesses: Item removal and item substitution, which together ensure that only relevant information is kept active in working memory (Ecker et al., 2014; Rac-Lubashevsky & Kessler, 2016a). Together, these processes allow working memory to alternate modes between flexible (when new information is encountered) and stable (when distractors are

¹We use the term 'item' to refer to an individual representation held in WM. 'Item' is thus synonymous with 'chunk' (Thalmann et al., 2019) and 'cognitive object' (Mathy & Feldman, 2012; Sewell et al., 2016) which denote the same concept. There is ongoing debate about whether items in WM are represented in *discrete slots* (items held with high precision in a number of discrete memory locations), allocation of *continuous resources* (items allocated limited resources in inverse proportion to the total number of items in WM), or some hybrid of the two frameworks (e.g., Bays & Husain, 2008; Donkin et al., 2013; Ma et al., 2014; Zhang & Luck, 2008). The models and general approach that we discuss in this paper are not committed to either architecture but could be used to test between the competing accounts (see 'Current directions' below).

encountered). This enables successful performance in dynamic environments in which distractions are common and the relevance of information frequently changes.

To date, the most detailed neurocomputational account of the gating mechanism controlling the trade-off between updating and maintenance is the prefrontal cortex-basal ganglia working memory (PBWM) model (Figure 5.1; Frank et al., 2001; Hazy et al., 2006; O'Reilly & Frank, 2006). In this model, gating is implemented via basal ganglia (BG)-thalamus-prefrontal cortex (PFC) circuits that control 'go/no-go' signalling. As illustrated in Figure 5.1, gate opening is controlled by a striatal 'go' signal which inhibits substantia nigra pars reticulata (SNr) and disinhibits the thalamus, which in turn excites PFC. This allows information to enter working memory and updating to occur. Gate closing² is controlled by a striatal 'no-go' signal, which inhibits external globus pallidus (GPe), disinhibits SNr, and inhibits the thalamus. This, in turn, inhibits PFC, which prevents working memory from being updated (Figure 5.1; 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. These circuits have also been implicated in updating value representations in reinforcement learning and value-based decision-making, suggesting a general neural mechanism for accomplishing information gating (Bledowski et al., 2010; Cools et al., 2007; Frank & O'Reilly, 2006; Hazy et al., 2006; Jocham et al., 2011; O'Reilly, 2006; Roth et al., 2005). The components of the PBWM model have received broad support from functional magnetic resonance imaging (fMRI) studies (Cools et al., 2007; Dahlin et al., 2008; Lewis et al., 2004; McNab & Klingberg, 2008; Murty et al., 2011; Roth et al., 2005; Tan et al., 2007). For example, activation in striatum and dorsolateral PFC has been widely reported in tasks broadly involving working memory updating (e.g., Dahlin et al., 2008; Lewis et al., 2004; McNab & Klingberg, 2008; Tan et al., 2007), while other work has localised activity specifically related to the updating and gating processes rather than other working memory processes. Roth et al. (2005) identified a frontoparietal network specifically involved in updating, while Murty et al. (2011) found selective engagement of SN/ventral tegmental area (VTA), caudate, dorsolateral PFC, and some areas of parietal cortex-related to the updating but not maintenance mode of working memory. Striatal dopamine-receptor-expressing neurons and dopamine-producing midbrain structures have also been implicated in working memory

² The PBWM model assumes that WM sits in the 'gate-closed'/maintenance mode by default. We note that 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 behavioural data (e.g., Rac-Lubashevsky & Kessler, 2016a, 2016b).

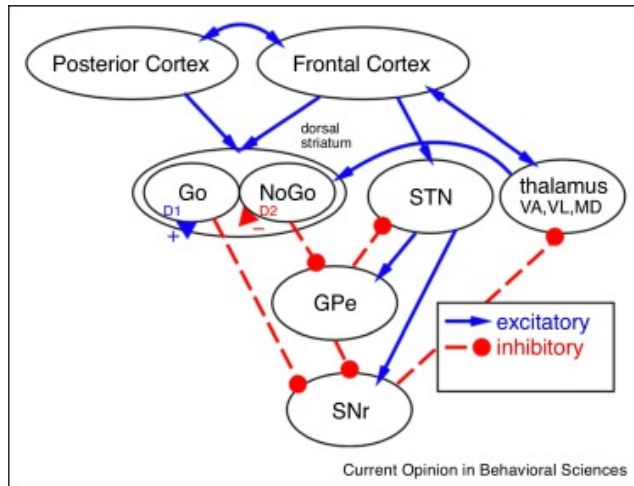


Figure 5.1. Illustration of the reference-back task. On each trial, participants indicate whether the presented letter is same or different from the letter in the most recent red frame. On reference (red frame) trials, participants must also update WM with the currently displayed letter. On comparison (blue frame) trials, participants make the same/different decision but do not update WM. Comparing behavioural outcomes (e.g. response time, error rate) between different trial types measures the cost of gate opening, gate closing, updating, and item substitution processes (see text for details). Explaining these behavioural phenomena via computational cognitive models and establishing further links to neural data is a key goal of current WM research. Adapted from Rac-Lubashevsky and Kessler (2016a) with permission.

updating (Cools et al., 2007; McNab & Klingberg, 2008; Murty et al., 2011), and the dynamic causal modelling suggests that BG plays a central role in gating information to PFC (van Schouwenburg, den et al., 2010). Moreover, a number of cortical areas (e.g., dorsolateral PFC, medial PFC, posterior parietal cortex) have been linked to the maintenance mode of working memory but not updating (Roth et al., 2005; Feredoes et al., 2011; D’Esposito & Postle, 2015; Owen et al., 2005). This is consistent with the idea that tonic dopamine activity in PFC controls the stability of working memory representations, whereas phasic dopamine release in the striatum trains the BG when to open the gate (via disinhibition of thalamus and PFC) to allow information into WM³ (O’Reilly & Frank, 2006).

³The 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 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

Overall, these findings show that working memory updating engages cortico-striatal circuitry involving BG, midbrain, and PFC structures broadly in line with the neurocomputational mechanisms of the PBWM model (Chatham & Badre, 2015; Hazy et al., 2006) and more general accounts of cognitive control (e.g., Braver, 2012). However, as will be discussed, recent work highlights that working memory also depends on several important subprocesses not accounted for in the PBWM and on neural substrates outside of the PBWM's BG-thalamus-PFC pathways. Modelling these processes and their neural basis is necessary to achieve a complete neurocomputational understanding of working memory.

This review discusses recent progress toward this goal. We focus on recent efforts to link brain measurements with behaviour on the reference-back task (Figure 5.2; Rac-Lubashevsky & Kessler, 2016a, 2016b), a working memory-based decision-making task that provides separate behavioural measures of gate opening and closing, as well as updating and substitution processes not accounted for in the PBWM. In doing so, we suggest that further progress can be made by applying the methods of model-based cognitive neuroscience (Forstmann & Wagenmakers, 2015; Turner et al., 2019), which links brain activity to behaviour via detailed computational models of cognitive and neural processes (Friston, 2009; Love, 2016; Schall, 2019). Model-based cognitive neuroscience generates detailed quantitative theories that span multiple levels of abstraction (e.g., behavioural, cognitive, neural). This provides greater constraint on theory and leads to more robust and detailed inferences. In particular, combining model-based approaches with developments in ultra-high field fMRI enables testing neurocomputational theories of working memory (such as the PBWM) with greater spatial and psychometric precision than has previously been possible. Applying these methods to the reference-back task promises a more detailed neurocomputational understanding of working memory than is currently available.

5.2 Measuring working memory subprocesses with the reference-back paradigm

Most laboratory tasks used to study working memory (e.g. n-back, delayed-match-to-sample) are designed to investigate the capacity and temporal properties of working memory but are unable to differentiate the contribution of working memory subprocesses to observed behaviour (Ecker et al., 2010; Nir-Cohen et al., 2020; Lewis-Peacock et al., 2018; Rac-Lubashevsky & Kessler, 2016a, 2016b; Roth et al., 2005). A recently developed exception is the *reference-back task* (Rac-Lubashevsky &

Kessler, 2016a, 2016b), which provides dissociable measures of core working memory subprocesses (gate opening, gate closing, updating, substitution) from behavioural choice-response time (RT) data.

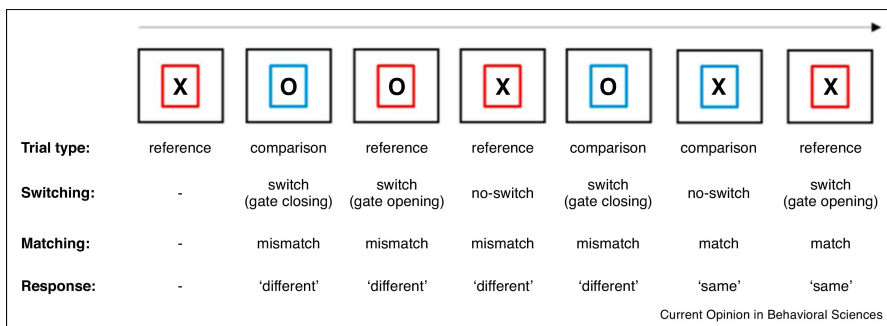


Figure 5.2. Illustration of the reference-back task. On each trial, participants indicate whether the presented letter is same or different from the letter in the most recent red frame. On reference (red frame) trials, participants must also update WM with the currently displayed letter. On comparison (blue frame) trials, participants make the same/different decision but do not update WM. Comparing behavioural outcomes (e.g., response time, error rate) between different trial types measures the cost of gate opening, gate closing, updating, and item substitution processes (see text for details). Explaining these behavioural phenomena via computational cognitive models and establishing further links to neural data is a key goal of current WM research. Adapted from Rac-Lubashevsky and Kessler (2016a) with permission.

To perform the reference-back, participants hold one of two stimuli (e.g. an ‘X’ or ‘O’) in working memory while deciding whether a series of probes matches the current item in working memory (Figure 5.2). On reference trials (indicated by a red frame around the stimulus), the participant must update working memory 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 working memory (the one appearing in the most recent red frame) without updating working memory. 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 behavioural measure of the cost of updating. By similar logic, switching from comparison to reference trials requires opening the working memory gate (to allow for updating), while switching from reference to comparison trials requires closing the working memory gate (to maintain the current contents). Gate opening is measured by comparing trials on which participants switch towards a reference trial to those where reference trials are repeated. Likewise, gate closing is measured by comparing trials on which participants switch towards 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 working memory.

The benchmark behavioural finding from the reference-back task is that trials requiring additional working memory processes tend to have slower RTs and/or more frequent errors than trials that do not require such processes (Jongkees, 2020; Nir-Cohen et al., 2020; Rac-Lubashevsky & Kessler, 2016a, 2016b, 2017, 2018, 2019). 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 working memory representations due to drawing attention/capacity away from the decision process) (Pearson et al., 2014). However, distinguishing these accounts requires detailed choice-RT models of the latent cognitive processes underlying memory-based decision-making (e.g. the highly successful evidence accumulation framework; Brown & Heathcote, 2008; Ratcliff, 1978), which are yet to be applied to the reference-back paradigm. Before discussing approaches to modelling the reference-back task, we first review recent efforts to identify the neural substrates of working memory subprocesses by correlating brain activity with behavioural measures derived from the reference-back.

5.3 Neural correlates of the reference-back task

As outlined above, there is broad consensus from neuroimaging supporting the role of BG, thalamus, and PFC in working memory gating as instantiated in the PBWM (O'Reilly & Frank, 2006). However, the neural basis of several core working memory subprocesses (e.g. gate closing, updating, substitution) is less clear. Recent work has begun to address this gap by linking behavioural measures derived from the reference-back with neurophysiological measures such as EEG and fMRI (Jongkees, 2020; Nir-Cohen et al., 2020; Rac-Lubashevsky & Kessler, 2018, 2019).

Two initial studies investigated EEG correlates of the reference-back task. Rac-Lubashevsky and Kessler (2018) found that gate closing was associated with increased theta power, a neural signature of cognitive control (Cavanagh & Frank, 2014; Cohen, 2014; Cohen & Donner, 2013), while gate opening and updating were associated with increased delta power, a signature of reactive (event-driven) control and action selection processes that engage in response to reward prediction errors (Cavanagh, 2015; Gulbinaite et al., 2014; Harmony, 2013). This suggests a functional role for delta and theta signals in the control of working memory consistent with 'go/no-go' signalling in the PBWM model (Chatham & Badre, 2015; Frank et al.,

2001; Hazy, Frank, O'Reilly, 2006). A follow-up study explored the role of the P3b EEG signal (a positive event-related potential that signals task-relevant events and peaks 300 ms after stimulus onset) in gating and updating (Rac-Lubashevsky & Kessler, 2019). P3b amplitude spiked depending on whether the stimulus matched the working memory reference item, implicating P3b in stimulus comparison/categorisation processes rather than updating per se. Greater negative activity (in an N2-like ERP component unrelated to the P3b) was found in anterior cortical regions on reference versus comparison trials. This signal has been implicated in controlled inhibition and action selection (Folstein & van Petten, 2008) and, in the context of the reference-back task, likely reflects a gate-opening or updating signal, consistent with the PBWM's assumption that reference trials trigger an update or 'go' signal to allow new information into working memory. This initial work demonstrates that neural signatures of specific updating and gating processes are detectable in EEG oscillatory signals that show activity broadly consistent with 'go/no-go' signalling in BG-thalamus-PFC pathways involved in working memory gating (Frank et al., 2001; Hazy, Frank, O'Reilly, 2006). However, the poor spatial resolution of EEG limits our ability to draw conclusions about the specific structures associated with each working memory subprocess.

Extending this work, Nir-Cohen et al. (2020) used 3T fMRI to identify neural substrates of working memory subprocesses using a modified reference-back with more complex face-morph stimuli. BG, frontoparietal cortex, and task-relevant sensory areas such as the visual cortex were involved in gate opening. Gate closing activated the parietal cortex, and substitution elicited activation in the left dorsolateral PFC and inferior parietal lobule. A whole-brain conjunction analysis revealed shared activity in the supplementary motor area for updating and substitution while updating and gating both activated the posterior parietal cortex. These results broadly agree with the PBWM model (Hazy, Frank, O'Reilly, 2006) and support the role of BG and PFC in controlling the flow of information into working memory and replacing old with new information. However, parietal cortex activation during gate closing is not predicted by the PBWM. This suggests that additional brain structures are involved in controlling working memory subprocesses and points to an opportunity to extend the PBWM to explain the neural basis of working memory subprocesses beyond gate opening.

Jongkees (2020) provided further evidence for the dopaminergic basis of working memory gating and updating processes by administering dopamine precursor L-tyrosine to young adults and comparing reference-back performance to a placebo-control condition. In the L-tyrosine condition, participants had fewer

variable gate opening times than in the placebo control condition, suggesting that the drug improved working memory performance for poor performers but impaired high performers. There was no effect on updating or gate closing, consistent with the role of striatal dopamine signals in opening the gate to working memory in line with the PBWM (Frank et al., 2001; Hazy, Frank, O'Reilly, 2006). Further indirect support for striatal dopamine involvement comes from a study linking event-based eye-blink rate (a proxy measure of striatal dopamine) to working memory updating in the reference-back task (Rac-Lubashevsky et al., 2017). However, follow-up work combining this approach with ultra-high field fMRI is needed to identify how activity in small subcortical structures, as well as layers in the cortex (e.g. striatum, GP, thalamus, PFC), is modulated by dopamine.

5.4 Current directions

The work reviewed above has taken important first steps toward identifying the neural substrates of working memory subprocesses beyond the BG-thalamus-PFC 'go/no-go' gating mechanism of the PBWM (Chatham & Badre, 2015; Hazy, Frank, O'Reilly, 2006). However, existing work has so far been limited to relating brain activity directly to the reference-back's behavioural measures rather than the latent cognitive processes that give rise to behaviour. Model-based approaches that link brain and behaviour via computational cognitive models offer numerous advantages over traditional statistical analyses of mean RT and error rate in understanding the cognitive and neural basis of working memory. For example, applying evidence accumulation models of choice-RT (e.g., Brown & Heathcote, 2008; Ratcliff, 1978) to reference-back data would reveal whether performance costs occur because working memory subprocesses add time outside of the decision stage (longer non-decision time), interfere with the decision process itself (reduced or noisier processing rate; Pearson et al., 2014), or induce strategic adjustments engaging top-down cognitive control (increased response caution). Decomposing behavioural effects (e.g. gating, updating costs) into a set of latent cognitive processes (e.g. accumulation rate, non-decision time, cognitive control of thresholds) rather than coarse behavioural-level summary statistics enables exploring the neural substrates of working memory in greater detail than is possible with traditional methods (de Hollander et al., 2016; Forstmann et al., 2011). This places stronger constraints on theory and ultimately produces more robust and detailed inferences about the latent processes that generate behaviour. Applying cognitive models to the reference-back holds great promise in this regard.

In its standard form, the reference-back paradigm ignores several important additional working memory processes. These include mechanisms that operate on information already active in working memory (Logie et al., 1994; Owen et al., 1997; Furst & Hitch, 2000), such as object selection and retrieval (Sewell et al., 2016), item-specific removal (Ecker et al., 2014; but see Kessler, 2018, for evidence of removal in the reference-back), and grouping and reorganisation operations (e.g., sorting items into alphabetical or chronological order, chunking or grouping items together to form a single accessible representation, changing the serial position of items; D'Esposito et al., 1999; Marshuetz, 2005; Nassar et al., 2018; van Dijck et al., 2013). These mechanisms support effective remembering by restructuring information into more memorable formats and ensuring only relevant information is maintained and retrieved from working memory. The standard reference-back also ignores phenomena associated with working memory's limited capacity (e.g., WM load/set-size effects; Collins & Frank, 2012; Donkin et al., 2013; McDougle & Collins, 2020; Sewell et al., 2014) and the temporal degradation (e.g., by decay or interference) of working memory representations (for a review, see Ricker et al., 2016). Analyses that do not account for these processes risk misattributing their effects to other processes, resulting in biased inferences.

Simple extensions to the reference-back task (e.g. using multiple-item working memory sets, inserting delays between the update cue and stimulus presentation), however, enable testing such effects alongside the gating and updating processes of the standard reference-back. For example, Verschooren et al. (2021) developed a modified reference-back paradigm where one among several items in long-term memory or perception is gated into working memory. This allows for comparing gating dynamics for perceptual versus long-term memory information. Similar multiple-item modifications can be used to investigate some of the working memory phenomena described above, including informing the ongoing debate about whether items in working memory are held in a small number of discrete high-precision slots (Donkin et al., 2013) or allocated capacity from a limited pool of continuous resources (Bays & Husain, 2008; Ma et al., 2014; Zhang & Luck, 2008). In discrete slots models, the fidelity of items in working memory only degrades once all memory slots are full (e.g. when $n > 4$). In continuous resource models, an item's fidelity is determined by its share of the available resources and, thus, should degrade in inverse proportion to the total number of items in WM⁴. Evidence accumulation

⁴Note, however, that continuous resource models can mimic discrete slots models. For example, if a resource pool has capacity to accommodate four items, then item fidelity may only begin to degrade once demands exceed capacity (i.e. when $n > 4$), thus producing similar predictions to a discrete slots model. Careful experimental design is needed in order to correctly attribute effects to capacity limitations (Navon, 1984).

models are well suited to test between these competing accounts (e.g. via accumulation rate parameters) as they can be used to assess the fidelity of working memory representations and measure capacity-sharing effects; Donkin et al., 2013; Eidels et al., 2010). Varying set sizes in the reference-back and assessing the impact on decision-making and working memory processes (as measured by cognitive models) could test between slots and resource architectures. Similarly, combining a multiple-item reference-back task with reinforcement learning (e.g. by reinforcing some items but not others) could shed light on the interplay between working memory and learning (e.g., Collins & Frank, 2012; McDougle & Collins, 2020) and the role of expected value in working memory-based decisions. Overall, we believe detailed choice-RT modelling will play an essential role in resolving these critical questions and explaining additional working memory phenomena captured by variants of the reference-back task.

Combining computational approaches with recent developments in ultra-high field fMRI (7T and higher) (e.g. increased resolution and better signal- and contrast-to-noise ratios) holds great promise for identifying activity in small subcortical structures (e.g., GP, SN, subthalamic nucleus, VTA; de Hollander et al., 2017; Trutti et al., 2019) and gaining a deeper understanding of their functional role in working memory than is currently available. For example, this would enable a stronger test of the so-called ‘third phase’ response of the PBWM model (O’Reilly & Frank, 2006), which evaluates the updating process via dopaminergic midbrain neurons that code reward prediction errors (Schultz, 1997). Under the PBWM, midbrain dopamine responses that train the BG when to update should no longer occur once updating-related task rules have been learned. This mechanism has proven difficult to verify with low-field strength fMRI (D’Ardenne et al., 2012; Yu et al., 2013); however, imaging reference-back performance with ultra-high field fMRI and linking neural measurements to cognitive model parameters would enable identifying these anatomical and functional mechanisms in greater detail and provide an additional constraint on cognitive models of working memory.

Specifically, when modelling two or more sources of data (e.g. fMRI and choice-RT) simultaneously, the power to detect joint effects (e.g. correlations between BOLD signal and cognitive model parameters) is determined by the signal-to-noise ratios of each data source. Increasing the signal-to-noise ratio of neural data (e.g. via 7T fMRI; de Hollander et al., 2017) reduces uncertainty throughout the model, as does including data from additional modalities (e.g., EEG + fMRI + behavioural; Turner et al., 2016)⁵. A further benefit is that connecting neural signals

⁵ This is particularly important for individual differences analyses, which rely on precise measurement at the individual level to accurately capture the variation between people.

to cognitive model parameters allows for selecting between cognitive models that make identical predictions at the level of choice-RT but differ in their internal dynamics (Ditterich, 2010; Forstmann et al., 2011; Hawkins et al., 2017; Schall, 2019). That is, different internal mechanisms can be titrated by evaluating which is most consistent with the additional structure provided by the neural data. Combining such approaches with the reference-back task has potential to shed light on other structures known to be involved in working memory (e.g. hippocampus; Norman & O'Reilly, 2003; O'Reilly & Norman, 2002; Weilbacher & Gluth, 2016), dopaminergic response evaluation (e.g. VTA; Bouarab et al., 2019; van Zessen et al., 2012), and cognitive control (e.g., anterior cingulate cortex; Shenhav et al., 2013), which are not yet accounted for in existing neurocomputational models. Linking state-of-the-art fMRI to the latent cognitive processes engaged by the reference-back would offer particular insight into the function of small dopamine-producing midbrain structures, with implications for understanding working memory impairments in a range of clinical disorders involving abnormal dopamine function (Huys et al., 2016). Overall, we believe that viewing the reference-back task through the lens of model-based cognitive neuroscience promises a more detailed understanding of the subprocesses that support working memory and their neural substrates.

5.5 Concluding remarks

This review discussed recent efforts to identify the neural basis of subprocesses that support working memory in the recently developed reference-back task. Current empirical work supports the idea that working memory gating is controlled by striatal 'go/no-go' signalling in BG-thalamus-PFC pathways. However, the neural substrates of several additional working memory subprocesses are yet to be established, pointing to a need for ultra-high field functional imaging combined with detailed computational cognitive modelling. Targets for future research include extending the reference-back task to account for additional working memory subprocesses (e.g. removal, selection, and reorganisation operations) and effects of working memory load and capacity (e.g. longer retrieval times, noisier working memory representations), as ignoring such processes leads to mis specified models and potentially biased inferences. Applying the methods of model-based cognitive neuroscience to the reference-back task would provide a major advance in understanding working memory at neural, cognitive, and behavioural levels. A comprehensive understanding of working memory subprocesses and their neural basis is within reach, with implications for both cognitive and clinical neuroscience.

