Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Behavioral Sciences

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

Anne C Trutti[1,2], Sam Verschooren[3], Birte U Forstmann[1] and Russell J Boag[1]

Working memory (WM) refers to a set of processes that makes task-relevant information accessible to higher-level cognitive processes. Recent work suggests WM is supported by a variety of information gating, updating, and removal processes, which ensure only task-relevant information occupies WM. Current neurocomputational theory suggests WM 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 WM. We review recent efforts to identify the neural basis of WM subprocesses using the recently developed reference-back task as a benchmark measure of WM subprocesses. Targets for future research using the methods of model-based cognitive neuroscience and novel extensions to the reference-back task are suggested.

Addresses
[1] Department of Psychology, University of Amsterdam, Amsterdam, The Netherlands
[2] Institute of Psychology, Leiden University, Leiden, The Netherlands
[3] Department of Experimental Clinical and Health Psychology, Ghent University, Ghent, Belgium

Corresponding author:
Trutti, Anne C (r.j.boag@uva.nl), Boag, Russell J (r.j.boag@uva.nl)

Current Opinion in Behavioral Sciences 2020, **38**:57–65

This review comes from a themed issue on **Computational cognitive neuroscience**

Edited by **Geoff Schoenbaum** and **Angela J Langdon**

https://doi.org/10.1016/j.cobeha.2020.10.002

Working memory and its subprocesses

Working memory (WM) 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 [1–3]. Working memory is extremely *capacity-limited*, with current research suggesting that between one and four items[4] can be maintained in an activated state in WM at a time [4–7]. This strict limit demands a high degree of control over WM content, such that 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or distracting information) and *flexibility* (i.e. keeping WM up-to-date with new relevant information and removing outdated information). This trade-off between stability and flexibility [8–11] is a core feature of executive control processes (e.g. cognitive control, conflict monitoring/resolution, task switching; [12]) and managing the trade-off strongly depends on the brain's dopamine systems [13•,14].

Prominent computational theories suggest that WM resolves the stability-flexibility trade-off by operating in two modes: An *updating* (gate-open) mode, which allows new information to enter WM, and a *maintenance* (gate-closed) mode, which prevents irrelevant and distracting information from interfering with the current contents of WM [15–22]. 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 WM [23,24••]. Together, these processes allow WM to alternate modes between *flexible* (when new information is encountered) and *stable* (when distractors are 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

[4] We use the term 'item' to refer to an individual representation held in WM. 'Item' is thus synonymous with 'chunk' [97] and 'cognitive object' [98,7] 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. Refs. [78,74,79,80]). 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 Section 'Current directions' below).

Figure 1

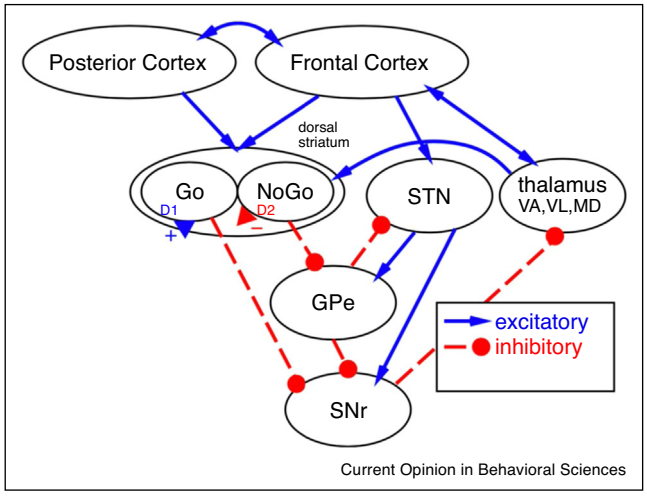

Illustration of the PBWM model. Gate opening is controlled by a striatal 'go' signal that inhibits SNr and disinhibits thalamus and PFC, enabling updating to occur. Gate closing is controlled by a striatal 'no-go' signal that inhibits GPe, disinhibits SNr, which inhibits thalamus and PFC, preventing updating. Extending this model to include additional structures implicated in WM and cognitive control (e.g. hippocampus, ventral tegmental area, anterior cingulate cortex) and their role in WM subprocesses beyond gate opening/closing is a key target for model-based cognitive neuroscience. Adapted from Hazy *et al.* [26] with permission.

cortex-basal ganglia WM (PBWM) model (Figure 1; [25–27]). 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 1, gate opening is controlled by a striatal 'go' signal which inhibits substantia nigra pars reticulata (SNr) and disinhibits thalamus, which in turn excites PFC. This allows information to enter WM and updating to occur. Gate closing[5] is controlled by a striatal 'no-go' signal which inhibits external globus pallidus (GPe), disinhibits SNr, and inhibits thalamus. This in turn inhibits PFC, which prevents WM from being updated (Figure 1; [26]). 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 ([16,28,29,26,20,30,22]).

The components of the PBWM model have received broad support from functional magnetic resonance imaging (fMRI) studies ([19,28,31–33,22,34]). For example, activation in striatum and dorsolateral PFC has been widely reported in tasks broadly involving WM updating (e.g. Refs. [31–34]), while other work has localized activity specifically related to the updating and gating processes rather than other WM processes. Roth *et al.* [22] identified a frontoparietal network specifically involved in updating, while Murty *et al.* [19] 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 WM. Striatal dopamine-receptor expressing neurons and dopamine-producing midbrain structures have also been implicated in WM updating [19,28,33], and dynamic causal modelling suggests that BG plays a central role in gating information to PFC [35]. Moreover, a number of cortical areas (e.g. dorsolateral PFC, medial PFC, posterior parietal cortex) have been linked to the maintenance mode of WM but not updating ([22,36–38]). This is consistent with the idea that tonic dopamine activity in PFC controls the stability of WM 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 WM[6] [27].

Overall, these findings show that WM updating engages cortico-striatal circuitry involving BG, midbrain, and PFC structures broadly in line with the neurocomputational mechanisms of the PBWM model [39,26] and more general accounts of cognitive control (e.g. Ref. [40]). However, as will be discussed, recent work highlights that WM 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 WM.

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 2; [24••,21•]), a WM-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

[5] 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. Refs. [24••,21•]).

[6] 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 [84]. 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.

Figure 2

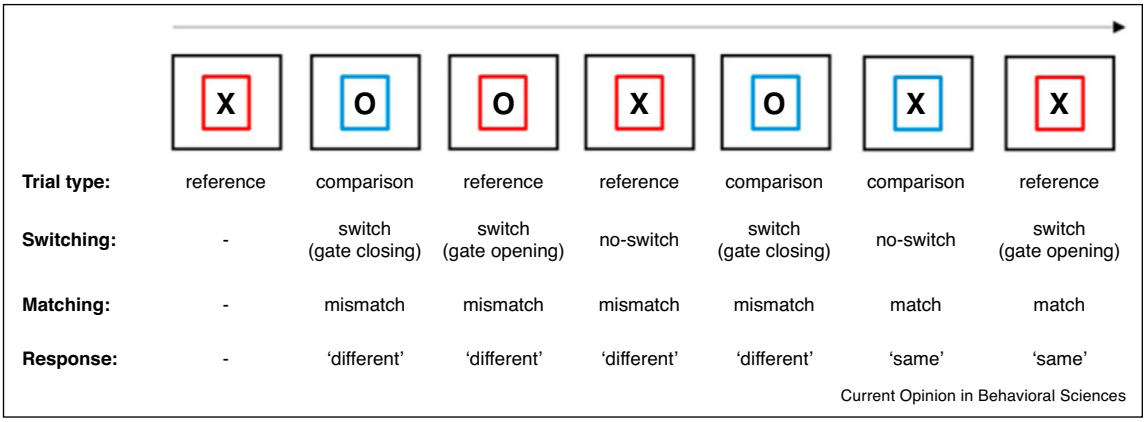

Illustration of the reference-back task. On each trial, participants indicate whether the presented letter is same or different from the letter in the most recent red frame. On reference (red frame) trials, participants must also update WM with the currently displayed letter. On comparison (blue frame) 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 [21•] with permission.

further progress can be made by applying the methods of *model-based cognitive neuroscience* [41,42•], which links brain activity to behaviour via detailed computational models of cognitive and neural processes [43–45]. 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 WM (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 WM than is currently available.

Measuring WM subprocesses with the reference-back paradigm

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 behaviour ([24••,46,47••,48•,21•,22]). A recently developed exception is the *reference-back* task [24••,21•], which provides dissociable measures of core WM subprocesses (gate opening, gate closing, updating, substitution) from behavioural choice-response time (RT) data.

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 (Figure 2). 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 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 behavioural 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 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 WM.

The benchmark behavioural 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 [24••,47••,49•,21•,50•,51•,52•]. 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) [53]. 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, [54,55]), 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 WM subprocesses by correlating brain activity with behavioural measures derived from the reference-back.

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 WM gating as instantiated in the PBWM [27]. However, the neural basis of several core WM 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 [47••,49•,51•,52•].

Two initial studies investigated EEG correlates of the reference-back task. Rac-Lubashevsky and Kessler [51•] found that gate closing was associated with increased theta power, a neural signature of cognitive control [56–58], 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 [59–61]. This suggests a functional role for delta and theta signals in the control of WM consistent with 'go/no-go' signalling in the PBWM model [25,39,26]. 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 [52•]. P3b amplitude spiked depending on whether the stimulus matched the WM 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 [62] 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 WM. 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 WM gating [25,26]. However, the poor spatial resolution of EEG limits our ability to draw conclusions about the specific structures associated with each WM subprocess.

Extending this work, Nir-Cohen *et al.* [47••] used 3T fMRI to identify neural substrates of WM subprocesses using a modified reference-back with more complex face-morph stimuli. BG, frontoparietal cortex, and task-relevant sensory areas such as visual cortex were involved in gate opening. Gate closing activated parietal cortex and substitution elicited activation in 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 [26] and support the role of BG and PFC in controlling the flow of information into WM 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 WM subprocesses and points to an opportunity to extend the PBWM to explain the neural basis of WM subprocesses beyond gate opening.

Jongkees [49•] provided further evidence for the dopaminergic basis of WM gating and updating processes by administering dopamine precursor L-tyrosine to young adults and comparing reference-back performance to a placebo-control group. The L-tyrosine group had less variable gate opening times than placebo controls, suggesting that the drug improved WM 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 WM in line with the PBWM [25,26]. Further indirect support for striatal dopamine involvement comes from a study linking event-based eye-blink rate (a proxy measure of striatal dopamine) to WM updating in the reference-back task [50•]. 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 cortex (e.g. striatum, GP, thalamus, PFC) is modulated by dopamine.

Current directions

The work reviewed above has taken important first steps toward identifying the neural substrates of WM subprocesses beyond the BG-thalamus-PFC 'go/no-go' gating mechanism of the PBWM [39,26]. 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 WM. For example, applying evidence accumulation models of choice-RT (e.g. Refs. [54,55]) to reference-back data would reveal whether performance costs 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; [53]), 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, nondecision time, cognitive control of thresholds) rather than coarse behavioural-level summary statistics enables exploring the neural substrates of WM in greater detail than is possible with traditional methods [63,64]. 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 WM processes. These include mechanisms that operate on information already active in WM [65–67], such as object selection and retrieval [7], item-specific removal ([23]; but see Ref. [68], for evidence of removal in the reference-back), and grouping and reorganization 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; [69–72]). These mechanisms support effective remembering by restructuring information into more memorable formats and ensuring only relevant information is maintained and retrieved from WM. The standard reference-back also ignores phenomena associated with WM's limited capacity (e.g. WM load/set-size effects; [73–75,7]) and the temporal degradation (e.g. by decay or interference) of WM representations (for a review, see Ref. [76]). 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 WM 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.* [77] developed a modified reference-back paradigm where one among several items in long-term memory or perception is gated into WM. 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 WM phenomena described above, including informing the ongoing debate about whether items in WM are held in a small number of discrete high-precision slots [74] or allocated capacity from a limited pool of continuous resources [78–80]. In discrete slots models, the fidelity of items in WM 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 WM[7] . Evidence accumulation 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 WM representations and measure capacity-sharing effects; [74,81]). Varying set size in the reference-back and assessing the effects on decision-making and WM 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 WM and learning (e.g. Refs. [73,75]) and the role of expected value in WM-based decisions. Overall, we believe that detailed choice-RT modelling will play an important role in resolving these important questions and in explaining additional WM 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; [82,83]) and gaining a deeper understanding of their functional role in WM than is currently available. For example, this would enable a stronger test of the so-called 'third phase' response of the PBWM model [27], which evaluates the updating process via dopaminergic midbrain neurons that code reward prediction errors [84]. 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 [85,86], 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 additional constraint on cognitive models of WM. 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

[7] Note, 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 [99].

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; [82]) reduces uncertainty throughout the model, as does including data from additional modalities (e.g. EEG + fMRI + behavioural; [87])[8]. A further benefit is that connecting neural signals 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 [45,64,88,89]. 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 WM (e.g. hippocampus; [90,91,92•]), dopaminergic response evaluation (e.g. VTA; [93,94]), and cognitive control (e.g. anterior cingulate cortex; [95]), 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 WM impairments in a range of clinical disorders involving abnormal dopamine function [96]. 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 WM and their neural substrates.

Concluding remarks

This review discussed recent efforts to identify the neural basis of subprocesses that support WM in the recently developed reference-back task. Current empirical work supports the idea that WM gating is controlled by striatal 'go/no-go' signalling in BG-thalamus-PFC pathways. However, the neural substrates of several additional WM 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 WM subprocesses (e.g. removal, selection, and reorganization operations) and effects of WM load and capacity (e.g. longer retrieval times, noisier WM 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 WM at neural, cognitive, and behavioural levels. A comprehensive understanding of WM subprocesses and their neural basis is within reach, with implications for both cognitive and clinical neuroscience.

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

Author contributions

RJB and ACT conducted the literature review and led the writing of the manuscript. All authors provided feedback at different stages, reviewed, edited, and revised the manuscript.

Conflict of interest statement

Nothing declared.

Acknowledgements

We thank Dr Yoav Kessler for making several suggestions that improved an earlier version of this manuscript. This work was supported by a grant from the Netherlands Organisation for Scientific Research (NWO; grant number 016.Vici.185.052; BUF).

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Cowan N: **Evolving conceptions of memory storage, selective attention, and their mutual constraints within the human information-processing system**. *Psychol Bull* 1988, **104**:163.

2. Daneman M, Carpenter PA: **Individual differences in working memory and reading**. *J Mem Lang* 1980, **19**:450.

3. Kyllonen PC, Christal RE: **Reasoning ability is (little more than) working-memory capacity?!** *Intelligence* 1990, **14**:389-433.

4. Cowan N: **The magical number 4 in short-term memory: a reconsideration of mental storage capacity**. *Behav Brain Sci* 2001, **24**:87-114.

5. Garavan H: **Serial attention within working memory**. *Mem Cogn* 1998, **26**:263-276.

6. Oberauer K: **Access to information in working memory: exploring the focus of attention**. *J Exp Psychol Learn Mem Cognit* 2002, **28**:411.

7. Sewell DK, Lilburn SD, Smith PL: **Object selection costs in visual working memory: a diffusion model analysis of the focus of attention**. *J Exp Psychol Learn Mem Cognit* 2014, **42**:1673.

8. Dreisbach G: **Mechanisms of cognitive control: the functional role of task rules**. *Cur Dir Psychol Sci* 2012, **21**:227-231.

9. Dreisbach G, Fröber K: **On how to be flexible (or not): modulation of the stability-flexibility balance**. *Curr Dir Psychol Sci* 2019, **28**:3-9.

10. Hommel B: **Between persistence and flexibility: the Yin and Yang of action control**. *Advances in Motivation Science*. Elsevier; 2015:33-67.

11. Dreisbach G, Müller J, Goschke T, Strobel A, Schulze K, Lesch KP, Brocke B: **Dopamine and cognitive control: the influence of spontaneous eyeblink rate and dopamine gene polymorphisms on perseveration and distractibility**. *Behav Neurosci* 2005, **119**:483.

12. Musslick S, Jang SJ, Shvartsman M, Shenhav A, Cohen JD: **Constraints associated with cognitive control and the stability-flexibility dilemma**. *Proc. 40th Annual Meeting of the Cognitive Science Society; Cognitive Science Society: Madison, WI: 2018:804-809*.

13. • Cools R: **Chemistry of the adaptive mind: lessons from dopamine**. *Neuron* 2019, **104**:113-131.
Reviews the dopaminergic basis of the stability-flexibility trade-off faced by cognitive systems. Provides an outline of chemical neuromodulatory mechanisms and their role in enabling flexible goal-directed cognition.

14. Cools R, D'Esposito M: **Inverted-U-shaped dopamine actions on human working memory and cognitive control**. *Biol Psychiatry* 2011, **69**:e113-e125.

15. Badre D: **Opening the gate to working memory**. *Proc Natl Acad Sci U S A* 2012, **109**:19878-19879.

16. Bledowski C, Kaiser J, Rahm B: **Basic operations in working memory: contributions from functional imaging studies**. *Behavl Brain Res* 2010, **214**:172-179.

17. Kessler Y, Oberauer K: **Working memory updating latency reflects the cost of switching between maintenance and updating modes of operation**. *J Exp Psychol Learn Mem Cognit* 2014, **40**:738.

18. Miller EK, Cohen JD: **An integrative theory of prefrontal cortex function**. *Annu Rev Neurosci* 2001, **24**:167-202.

19. Murty VP, Sambataro F, Radulescu E, Altamura M, Iudicello J, Zoltick B, Weinberger DR, Goldberg TE, Mattay VS: **Selective updating of working memory content modulates meso-cortico-striatal activity**. *Neuroimage* 2011, **57**:1264-1272.

20. O'Reilly RC: **Biologically based computational models of high-level cognition**. *Science* 2006, **314**:91-94.

21. • Rac-Lubashevsky R, Kessler Y: **Decomposing the n-back task: an individual differences study using the reference-back paradigm**. *Neuropsychologia* 2016, **90**:190-199.

Explored individual differences in WM subprocesses measured by the reference-back task and compared reference-back performance to several variants of commonly used WM tasks. The reference-back task provided measures of how gate opening, gate closing, and updating each contribute to individual task performance. Controlling the contents of WM is shown to be a primary source of individual differences in WM-based decision making.

22. Roth JK, Serences JT, Courtney SM: **Neural system for controlling the contents of object working memory in humans**. *Cereb Cortex* 2006, **16**:1595-1603.

23. Ecker UK, Oberauer K, Lewandowsky S: **Working memory updating involves item-specific removal**. *J Mem Lang* 2014, **74**:1-15.

24. •• Rac-Lubashevsky R, Kessler Y: **Dissociating working memory updating and automatic updating: the reference-back paradigm**. *J Exp Psychol Learn Mem Cognit* 2016, **42**:951.

Introduced the reference-back paradigm. The experiments in this study provided initial behavioural evidence for a gating mechanism that separates WM from long-term memory.

25. Frank MJ, Loughry B, O'Reilly RC: **Interactions between frontal cortex and basal ganglia in working memory: a computational model**. *Cogn Affect Behav Neurosci* 2001, **1**:137-160.

26. Hazy TE, Frank MJ, O'Reilly RC: **Banishing the homunculus: making working memory work**. *Neuroscience* 2006, **139**:105-118.

27. O'Reilly RC, Frank MJ: **Making working memory work: a computational model of learning in the prefrontal cortex and basal ganglia**. *Neural Comput* 2006, **18**:283-328.

28. Cools R, Sheridan M, Jacobs E, D'Esposito M: **Impulsive personality predicts dopamine-dependent changes in frontostriatal activity during component processes of working memory**. *J Neurosci* 2007, **27**:5506-5514.

29. Frank MJ, O'Reilly RC: **A mechanistic account of striatal dopamine function in human cognition: psychopharmacological studies with cabergoline and haloperidol**. *Behav Neurosci* 2006, **120**:497-517 http://dx.doi.org/10.1037/0735-7044.120.3.497.

30. Jocham G, Klein TA, Ullsperger M: **Dopamine-mediated reinforcement learning signals in the striatum and ventromedial prefrontal cortex underlie value-based choices**. *J Neurosci* 2011, **31**:1606-1613.

31. Dahlin E, Neely AS, Larsson A, Bäckman L, Nyberg L: **Transfer of learning after updating training mediated by the striatum**. *Science* 2008, **320**:1510-1512.

32. Lewis SJ, Dove A, Robbins TW, Barker RA, Owen AM: **Striatal contributions to working memory: a functional magnetic resonance imaging study in humans**. *Eur J Neurosci* 2004, **19**:755-760.

33. McNab F, Klingberg T: **Prefrontal cortex and basal ganglia control access to working memory**. *Nat Neurosci* 2008, **11**:103-107.

34. Tan HY, Chen Q, Goldberg TE, Mattay VS, Meyer-Lindenberg A, Weinberger DR, Callicott JH: **Catechol-O-methyltransferase Val158Met modulation of prefrontal–parietal–striatal brain systems during arithmetic and temporal transformations in working memory**. *J Neurosci* 2007, **27**:13393-13401.

35. van Schouwenburg MR, den Ouden HE, Cools R: **The human basal ganglia modulate frontal-posterior connectivity during attention shifting**. *J Neurosci* 2010, **30**:9910-9918.

36. Feredoes E, Heinen K, Weiskopf N, Ruff C, Driver J: **Causal evidence for frontal involvement in memory target maintenance by posterior brain areas during distracter interference of visual working memory**. *Proc Natl Acad Sci U S A* 2011, **108**:17510-17515.

37. D'Esposito M, Postle BR: **The cognitive neuroscience of working memory**. *Annu Rev Psychol* 2015, **66**:115-142.

38. Owen AM, McMillan KM, Laird AR, Bullmore E: **N-back working memory paradigm: a meta-analysis of normative functional neuroimaging studies**. *Hum Brain Mapp* 2005, **25**:46-59.

39. Chatham CH, Badre D: **Multiple gates on working memory**. *Curr Opin Behav Sci* 2015, **1**:23-31.

40. Braver TS: **The variable nature of cognitive control: a dual mechanisms framework**. *Trends Cogn Sci* 2012, **16**:106-113.

41. Forstmann BU, Wagenmakers EJ (Eds): *An Introduction to Model-based Cognitive Neuroscience*. New York: Springer; 2015:139-156.

42. • Turner BM, Palestro JJ, Miletić S, Forstmann BU: **Advances in techniques for imposing reciprocity in brain-behavior relations**. *Neurosci Biobehav Rev* 2019, **102**:327-336.

Summarises recent advancements in approaches to simultaneously modelling brain-behaviour relationships using the joint modelling approach central to model-based cognitive neuroscience. Multiple neural data streams (e.g. EEG, fMRI) can be used to inform latent cognitive model parameters, thus providing a better mechanistic understanding of the neural basis of latent cognitive mechanisms that give rise to observed behaviour.

43. Friston KJ: **Modalities, modes, and models in functional neuroimaging**. *Science* 2009, **326**:399-403.

44. Love BC: **Cognitive models as bridge between brain and behavior**. *Trends Cogn Sci* 2016, **20**:247-248.

45. Schall JD: **Accumulators, neurons, and response time**. *Trends Neurosci* 2019.

46. Ecker UK, Lewandowsky S, Oberauer K, Chee AE: **The components of working memory updating: an experimental decomposition and individual differences**. *J Expl Psychol Learn Mem Cogn* 2010, **36**:170.

47. •• Nir-Cohen G, Kessler Y, Egner T: **Distinct neural substrates for opening and closing the gate from perception to working memory**. *bioRxiv* 2019 http://dx.doi.org/10.1101/853630.

Used 3T fMRI to identify neural substrates of WM subprocesses using the reference-back paradigm. This work identifies brain regions involved in gate opening, gate closing, updating, and substitution processes, broadly consistent with the PBWM model in which BG-thalamus-PFC pathways control the flow of information into and out of WM.

48. • Lewis-Peacock JA, Kessler Y, Oberauer K: **The removal of information from working memory**. *Ann N Y Acad Sci* 2018, **1424**:33-44.

Reviews behavioural and neural evidence for a selective removal process that is engaged to remove outdated information from WM. The removal mechanism serves to limit working memory load by ensuring only goal-relevant information is maintained in WM. The ways in which selective removal is distinct from temporal decay and interference are also discussed.

49. • Jongkees BJ: **Baseline-dependent effect of dopamine's precursor L-tyrosine on working memory gating but not updating**. *Cogn Affect Behav Neurosci* 2020:1-15.

Investigated the dopaminergic basis of WM gating and updating processes by administering dopamine precursor L-tyrosine to young adults

and comparing reference-back performance to a placebo-control group. Updating and gating were differentially affected by the dopaminergic manipulation, highlighting the importance of the brain's dopamine systems for controlling WM.

50. • Rac-Lubashevsky R, Slagter HA, Kessler Y: **Tracking real-time changes in working memory updating and gating with the event-based eye-blink rate**. *Sci Rep* 2017, **7**:1-9.
Linked event-based eye-blink rate (a proxy for striatal dopamine activity) to reference-back task performance to examine how gating and updating facilitate flexible WM updating. The relation of event-based eye-blink rate to striatal dopamine provides indirect evidence in support of the striatal 'go/no-go' updating signal proposed by neurocomputational models of WM like the PBWM.

51. • Rac-Lubashevsky R, Kessler Y: **Oscillatory correlates of control over working memory gating and updating: an EEG study using the reference-back paradigm**. *J Cogn Neurosci* 2018, **30**:1870-1882.
Linked brain measurements (EEG) with behaviour on the reference-back task. Gating and updating were associated with EEG signatures of cognitive control and action selection processes. Demonstrates that neural signatures of specific updating and gating processes are detectable in EEG oscillatory signals that show activity consistent with 'go/no-go' signalling in BG-thalamus-PFC pathways, as proposed by prominent neurocomputational theories of WM updating.

52. • Rac-Lubashevsky R, Kessler Y: **Revisiting the relationship between the P3b and working memory updating**. *Biol Psychol* 2019, **148**:107769.
Used EEG with the reference-back task to explore the role of the P3b oscilatory signal in WM updating. P3b amplitude was related to stimulus comparison/categorisation but not to updating itself. Suggests that P3b is a neural signature of a goal-directed target identification mechanism that improves WM-based decision making in line with task goals.

53. Pearson B, Raškevičius J, Bays PM, Pertzov Y, Husain M: **Working memory retrieval as a decision process**. *J Vision* 2014, **14**:2.

54. Brown SD, Heathcote A: **The simplest complete model of choice response time: linear ballistic accumulation**. *Cogn Psychol* 2008, **57**:153-178.

55. Ratcliff R: **A theory of memory retrieval**. *Psychol Rev* 1978, **85**:59.

56. Cavanagh JF, Frank MJ: **Frontal theta as a mechanism for cognitive control**. *Trends Cogn Sci* 2014, **18**:414-421.

57. Cohen MX: **A neural microcircuit for cognitive conflict detection and signaling**. *Trends Neurosci* 2014, **37**:480-490.

58. Cohen MX, Donner TH: **Midfrontal conflict-related theta-band power reflects neural oscillations that predict behavior**. *J Neurophysiol* 2013, **110**:2752-2763.

59. Cavanagh JF: **Cortical delta activity reflects reward prediction error and related behavioral adjustments, but at different times**. *Neuroimage* 2015, **110**:205-216.

60. Gulbinaite R, van Rijn H, Cohen MX: **Fronto-parietal network oscillations reveal relationship between working memory capacity and cognitive control**. *Front Hum Neurosci* 2014, **8**:761.

61. Harmony T: **The functional significance of delta oscillations in cognitive processing**. *Front Integr Neurosci* 2013, **7**:83.

62. Folstein JR, van Petten C: **Influence of cognitive control and mismatch on the N2 component of the ERP: a review**. *Psychophysiology* 2008, **45**:152-170.

63. de Hollander G, Forstmann BU, Brown SD: **Different ways of linking behavioral and neural data via computational cognitive models**. *Biol Psychiatry Cogn Neurosci Neuroimaging* 2016, **1**:101-109.

64. Forstmann BU, Wagenmakers EJ, Eichele T, Brown S, Serences JT: **Reciprocal relations between cognitive neuroscience and formal cognitive models: opposites attract?** *Trends Cogn Sci* 2011, **15**:272-279.

65. Logie RH, Gilhooly KJ, Wynn V: **Counting on working memory in arithmetic problem solving**. *Mem Cognit* 1994, **22**:395-410.

66. Owen AM: **The functional organization of working memory processes within human lateral frontal cortex: the contribution of functional neuroimaging**. *Eur J Neurosci* 1997, **9**:1329-1339.

67. Fürst AJ, Hitch GJ: **Separate roles for executive and phonological components of working memory in mental arithmetic**. *Mem Cognit* 2000, **28**:774-782.

68. Kessler Y: **N-2 repetition leads to a cost within working memory and a benefit outside it**. *Ann N Y Acad Sci* 2018, **1424**:268-277.

69. D'Esposito M, Postle BR, Ballard D, Lease J: **Maintenance versus manipulation of information held in working memory: an event-related fMRI study**. *Brain Cognit* 1999, **41**:66-86.

70. Marshuetz C: **Order information in working memory: an integrative review of evidence from brain and behavior**. *Psychol Bull* 2005, **131**:323.

71. Nassar MR, Helmers JC, Frank MJ: **Chunking as a rational strategy for lossy data compression in visual working memory**. *Psychol Rev* 2018, **125**:486.

72. van Dijck JP, Abrahamse EL, Majerus S, Fias W: **Spatial attention interacts with serial-order retrieval from verbal working memory**. *Psychol Sci* 2013, **24**:1854-1859.

73. Collins AG, Frank MJ: **How much of reinforcement learning is working memory, not reinforcement learning? A behavioral, computational, and neurogenetic analysis**. *Eur J Neurosci* 2012, **35**:1024-1035.

74. Donkin C, Nosofsky RM, Gold JM, Shiffrin RM: **Discrete-slots models of visual working-memory response times**. *Psychol Rev* 2013, **120**:873.

75. McDougle SD, Collins AG: **Modeling the influence of working memory, reinforcement, and action uncertainty on reaction time and choice during instrumental learning**. *Psychon Bull Rev* 2020:1-20.

76. Ricker TJ, Vergauwe E, Cowan N: **Decay theory of immediate memory: from Brown (1958) to today (2014)**. *Q J Exp Psychol* 2016, **69**:1969-1995.

77. Verschooren S, Kessler Y, Egner T: **Evidence for a single mechanism gating perceptual and long-term memory information into working memory**. *PsyArXiv* 2020:0-33 http://dx.doi.org/10.31234/osf.io/3m7up.

78. Bays PM, Husain M: **Dynamic shifts of limited working memory resources in human vision**. *Science* 2008, **321**:851-854.

79. Ma WJ, Husain M, Bays PM: **Changing concepts of working memory**. *Nat Neurosci* 2014, **17**:347.

80. Zhang W, Luck SJ: **Discrete fixed-resolution representations in visual working memory**. *Nature* 2008, **453**:233-235.

81. Eidels A, Donkin C, Brown SD, Heathcote A: **Converging measures of workload capacity**. *Psychon Bull Rev* 2010, **17**:763-771.

82. de Hollander G, Keuken MC, van der Zwaag W, Forstmann BU, Trampel R: **Comparing functional MRI protocols for small, iron-rich basal ganglia nuclei such as the subthalamic nucleus at 7 T and 3 T**. *Hum Brain Mapp* 2017, **38**:3226-3248 http://dx.doi.org/10.1002/hbm.23586.

83. Trutti AC, Mulder MJ, Hommel B, Forstmann BU: **Functional neuroanatomical review of the ventral tegmental area**. *Neuroimage* 2019, **191**:258-268.

84. Schultz W, Dayan P, Montague PR: **A neural substrate of prediction and reward**. *Science* 1997, **275**:1593-1599.

85. D'Ardenne K, Eshel N, Luka J, Lenartowicz A, Nystrom LE, Cohen JD: **Role of prefrontal cortex and the midbrain dopamine system in working memory updating**. *Proc Natl Acad Sci U S A* 2012, **109**:19900-19909.

86. Yu Y, FitzGerald TH, Friston KJ: **Working memory and anticipatory set modulate midbrain and putamen activity**. *J Neurosci* 2013, **33**:14040-14047.

87. Turner BM, Rodriguez CA, Norcia TM, McClure SM, Steyvers M: **Why more is better: simultaneous modeling of EEG, fMRI, and behavioral data**. *Neuroimage* 2016, **128**:96-115.

88. Ditterich J: **A comparison between mechanisms of multi-alternative perceptual decision making: ability to explain human behavior, predictions for neurophysiology, and relationship with decision theory**. *Front Neurosci* 2010, **4**:184.

89. Hawkins GE, Mittner M, Forstmann BU, Heathcote A: **On the efficiency of neurally-informed cognitive models to identify latent cognitive states**. *J Math Psychol* 2017, **76**:142-155 http://dx.doi.org/10.1016/j.jmp.2016.06.007.

90. Norman KA, O'Reilly RC: **Modeling hippocampal and neocortical contributions to recognition memory: a complementary-learning-systems approach**. *Psychol Rev* 2003, **110**:611.

91. O'Reilly RC, Norman KA: **Hippocampal and neocortical contributions to memory: advances in the complementary learning systems framework**. *Trends Cogn Sci* 2002, **6**:505-510.

92. • Weilbächer RA, Gluth S: **The interplay of hippocampus and ventromedial prefrontal cortex in memory-based decision making**. *Brain Sci* 2017, **7**:4.
Reviews neuroimaging research on the links between brain and behaviour in episodic memory and value-based decision making, with a focus on interactions between hippocampus and ventromedial prefrontal cortex. Provides several interesting suggestions for obtaining a more detailed understanding of brain-behaviour relationships in memory-based decision making.

93. Bouarab C, Polter AM, Thompson B: **VTA GABA neurons at the interface of stress and reward**. *Front Neural Circuits* 2019, **13**:78.

94. van Zessen R, Phillips JL, Budygin EA, Stuber GD: **Activation of VTA GABA neurons disrupts reward consumption**. *Neuron* 2012, **73**:1184-1194.

95. Shenhav A, Botvinick MM, Cohen JD: **The expected value of control: an integrative theory of anterior cingulate cortex function**. *Neuron* 2013, **79**:217-240.

96. Huys QJ, Maia TV, Frank MJ: **Computational psychiatry as a bridge from neuroscience to clinical applications**. *Nat Neurosci* 2016, **19**:404.

97. Thalmann M, Souza AS, Oberauer K: **How does chunking help working memory?** *J Exp Psychol Learn Mem Cognit* 2019, **45**:37.

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

99. Navon D: **Resources—a theoretical soup stone?** *Psychol Rev* 1984, **91**:216.