# Investigating working memory updating processes of the human subcortex using 7T MRI

Anne C. Trutti[1,2], Zsuzsika Sjoerds[2], Russell J. Boag[1], Solenn L.Y. Walstra[1], Steven Miletić[1,2], Scott S.J. Isherwood[1], Pierre-Louis Bazin[3], Bernhard Hommel[4], Sarah Habli[5], Desmond H.Y. Tse[5], Asta K. Håberg[5], and Birte Forstmann[1]

[1] Integrative Model-Based Neuroscience Research Unit, University of Amsterdam, Amsterdam, The Netherlands.
[2] Cognitive Psychology Unit, Institute of Psychology & Leiden Institute for Brain and Cognition, Leiden University, Leiden, Netherlands
[3] Full brain picture Analytics, Leiden, The Netherlands.
[4] Department of Psychology, Shandong Normal University, Jinan, China.
[5] Norwegian University of Science and Technology, Trondheim, Norway

## Abstract

A growing body of research suggests that dopamine is involved in working memory updating and that the striatum takes up a critical role in the subprocess of working memory gating (Braver & Cohen, 2000; Cools & D'Esposito, 2011; D'Ardenne et al., 2012; Jongkees, 2020). In this study, we investigated subcortical–in particular, possible dopaminergic–involvement in working memory updating subprocesses using the reference-back task and ultra-high field 7 Tesla fMRI. Using a scanning protocol optimized for BOLD-sensitivity in the subcortex, we found no evidence of subcortical activation during working memory gate opening, predominantly activations in frontoparietal network regions, which challenges the idea of a striatal gating mechanism. However, during gate closing, subcortical activation was observed. Furthermore, a ready-to-update mode demonstrated large-spread subcortical activation, including basal ganglia nuclei, suggesting that the basal ganglia are engaged in general updating processes rather than specifically controlling the working memory gate. Moreover, substituting new information into working memory elicited activation in dopamine-producing midbrain regions along with the striatum, thalamus, and prefrontal cortex, indicating engagement of the basal ganglia-thalamo-cortical loop possibly driven by (potential) dopaminergic activity. These findings expand our understanding of subcortical regions involved in working memory updating, shifting the focus from gate opening to substitution as a midbrain-driven updating process.

## Submission information

**Data availability statement**

The analyzed data is part of a larger collection that will be published in a public repository when the entire dataset is ready for publication.

**Code availability statement**

All code used for preprocessing and analyses of the data acquired in this study will be made available at https://github.com/ACTrutti/WM-updating-subcortex after the publication of this study.

**Funding**
This work was supported by grants from the European Research Council to B. U. F. (8674750) and B. H. (ERC-2015-AdG- 694722), and a Vici grant from the Netherlands Organization for Scientific Research to B. U. F. (016.Vici.185.052). Funding sources were not involved in study design, data collection, and interpretation, or the decision to submit the work for publication.

**Competing interests**
The authors declare that no competing interests exist.

**Impact statement**
The study reveals subcortical contributions to several working memory updating subprocesses, refines the idea of a striatal gating mechanism and provides initial evidence for a midbrain-driven basal-ganglia-thalamo-cortical loop working memory substitution process.

**Author contributions**
**Anne C. Trutti**: Conceptualization, Formal analysis, Methodology, Investigation, Writing – Original draft preparation, Writing – Review & editing, Visualization, **Zsuzsika Sjoerds**: Conceptualization, Methodology, Supervision, Writing – Review & editing, **Russell J. Boag**: Conceptualization, Formal analysis, Supervision, Writing – Original draft, Writing – Review & editing, **Solenn L. Y. Walstra**: Conceptualization, Methodology, Writing – Review & editing, **Steven Miletić**: Methodology, Software, Data curation, Formal analysis, Investigation, Writing – Review & editing, Visualization, **Pierre-Louis Bazin**: Investigation, Methodology, Software, Writing – Review & editing, **Scott J. S. Isherwood**: Data curation, Investigation, Methodology, **Bernhard Hommel**: Conceptualization, Investigation, Resources, Funding acquisition, Writing – Review & editing**, Sarah Habli**: Investigation, **Desmond H. Y. Tse**: Methodology, Investigation, Software, **Asta K. Håberg**: Resources, Funding acquisition, Project administration, Writing – Review & editing, **Birte U. Forstmann**: Conceptualization, Investigation, Resources, Project administration, Funding acquisition, Writing – Review & editing.

**Acknowledgments**
We would like to express our gratitude to Pål Erik Goa for supporting this study by facilitating data acquisition. We also thank Niek Stevenson, Roel van Dooren, and Bryant Jongkees for their valuable contributions to preceding work and for stimulating discussions on the reference-back paradigm.

# 1 Introduction

Dopamine exerts a profound neuromodulatory impact on neural activity and is capable of shaping behavior at a large scale. In particular, dopamine plays a central role in cognitive control, and leading theories detail how dopamine networks influence cognitive functions (Cools & D'Esposito, 2011; Durstewitz & Seamans, 2008; Ott & Nieder, 2019). These theories assign key roles to the mesocortical and nigrostriatal dopaminergic pathways, which originate in the midbrain and project to the prefrontal cortex (PFC) and cingulate (for the mesocortical pathway), and the striatum (for the nigrostriatal pathway). These pathways are believed to promote either cognitive stability (protection from distractors) or flexibility (ability to incorporate/adapt to new information), respectively (Armbruster et al., 2012; Cools, 2006; Cools & D'Esposito, 2011; Durstewitz & Seamans, 2008; Goschke & Bolte, 2014). This balance is known as the stability—flexibility trade-off (Dreisbach, 2012; Dreisbach & Fröber, 2019; Hommel, 2015), influencing cognitive functions like working memory. For instance, maintaining stable working memory representations relies on PFC activity, the effectiveness of which is closely tied to cortical dopamine levels controlled by the mesocortical pathway (Durstewitz & Seamans, 2008). By contrast, the ability to flexibly update working memory representations is thought to be indirectly regulated by dopamine through the basal ganglia, implicating mesolimbic and nigrostriatal pathways (Chatham & Badre, 2015; Cohen et al., 2002). Several studies have provided initial support for working memory updating via the basal ganglia (Chatham & Badre, 2015; Murty et al., 2011; Nir-Cohen et al., 2020, 2023). Even though the processes involved in allowing new information into working memory (i.e., updating) depend on dopamine (Cools & D'Esposito, 2011; Goldman-Rakic, 1995), the neural basis of working memory updating remains unclear.

General working memory functioning has been robustly associated with the fronto-parietal network (FPN), which consists of the dorsolateral and medial PFC (dl/mPFC) and the posterior parietal cortex (PPC) (Jacob & Nieder, 2014; Mehta et al., 2000; Nir-Cohen et al., 2020, 2023; Vallentin et al., 2012). Additionally, several studies have implicated the basal ganglia in working memory encoding and maintenance (Bedwell et al., 2005; Chang et al., 2007), updating (Chatham & Badre, 2015; Murty et al., 2011), and gating (Nir-Cohen et al., 2020). However, the exact neural processes involved in the updating of working memory, especially those involving the subcortical structures, are still not fully understood. Much of this uncertainty can be attributed to the relatively low signal-to-noise ratios of functional magnetic resonance imaging (fMRI) scans within the subcortex, as well as the inherent biophysical characteristics of subcortical nuclei, such as the presence of iron. These factors pose challenges when using conventional fMRI protocols to investigate the subcortex. In this study, we addressed this gap by conducting ultra-high field 7 Tesla (T) fMRI to shed light on the subcortical structures thought to underlie working memory updating subprocesses.

Possible mechanisms behind selectively gating new information into working memory, i.e., updating, are captured in the *prefrontal-cortex basal-ganglia working memory* (PBWM) model (Frank et al., 2001; Hazy et al., 2007; O'Reilly & Frank, 2006). Central to the PBWM model is its proposition of a dynamic gating mechanism, where activation of the striatum acts as the pivotal initiator to open the gate to working memory–resulting in disinhibition of the thalamus through

inhibition of the substantia nigra pars reticulata (SNr)–which in turn modulates the stability of working memory representations in PFC.

Initial support for the PBWM model has been found in studies employing the reference-back paradigm (Rac-Lubashevsky & Kessler, 2016a) in combination with EEG (Rac-Lubashevsky & Kessler, 2018), dopaminergic manipulations (Jongkees, 2020), and fMRI (Nir-Cohen et al., 2020). The reference-back task is designed to identify the behavioral signatures of several subprocesses involved in working memory updating. These include working memory gate opening and closing, substituting new items into working memory, and a contrast between updating and maintenance modes (indicating readiness to either update or shield the contents of working memory, respectively). These processes are derived by presenting a stream of reference and comparison stimuli, which require updating and maintenance, respectively. Gate opening represents the active process of switching from a maintenance mode to a mode that *facilitates* updating. Thus, opening the gate takes place when cortical cell assemblies are prepared for new information to enter working memory in trials that may call for updating. Similarly, gate closing occurs when information needs to be maintained, and thus, the working memory gate should be closed and a maintenance mode activated in order to avoid interference. Substitution represents the process of replacing old information with new information and thus represents the *actual act* of updating working memory. The updating mode reflects an open-gate state (successive reference trials) independent of whether updating of information is required. These subprocesses are often conflated in working memory tasks that tap into the PBWM model, such as the standard *n*-back task, task-switching paradigms, and the AX-CPT task (Ecker et al., 2010; Kessler, 2018; Lewis-Peacock, Kessler, & Oberauer, 2018). In contrast, the reference-back task orthogonally differentiates these subprocesses–specifically isolating the crucial process of substituting a single stimulus in working memory from other facilitating subprocesses–providing a more precise mapping to brain regions, particularly in the deeper subcortical areas.

Nir-Cohen et al. (2020) found striatal activity associated with gate opening and substitution with fMRI at 3 T. Moreover, they observed no evidence of subcortical involvement in the other aspects of working memory updating, like gate closing. This implies that distinct neural mechanisms control gate opening and gate closing. Nir-Cohen et al. suggested the involvement of the mesocortical pathway as a possible explanation for the lack of striatal activity during gate closing in line with the *dual-state theory* (Durstewitz & Seamans, 2008; O'Reilly, 2006), which postulates that mesocortical dopaminergic neurons of the ventral tegmental area (VTA) modify PFC neuronal firing to influence the stability of working memory representations. This implies the possibility of a direct dopaminergic mechanism originating in the midbrain governing the switch from updating mode to maintenance mode in gate closing. Also, in the PBWM model, dopamine may take a key role in the selective updating of working memory as the striatal pathways of the basal ganglia are abundant in dopamine receptors, enabling dopamine to affect basal ganglia processes critically. The neural origin of the striatal dopamine is found in the midbrain's dopaminergic cell populations, VTA, and SN. In fact, a revised version of the PBWM model for Parkinson's disease patients (Moustafa, Sherman, & Frank, 2008) accounts for the interaction of dopamine with working memory by including the SN pars compacta (SNc), emphasizing the role of dopamine in working memory updating.

Since dopaminergic functioning in the brain is almost exclusively based on the activity of dopaminergic cell assemblies in the midbrain, the VTA, and SN–specifically SNc, these areas warrant special attention in explorations of dopamine-related processes. Initial fMRI evidence suggests a link between working memory and midbrain activity. There were indications of the involvement of the VTA and SN in the updating of context-related information (D'Ardenne et al., 2012). Additionally, the combined processing of information removal and updating (Murty et al., 2011) also implicated these midbrain structures. Moreover, Hazy, Frank, and O'Reilly (2006) provide their biologically based cognitive architecture onto the PBWM model and, in fact, suggest that dopaminergic signals arising from midbrain nuclei, specifically the VTA and SNc, modulate learning when striatal gating to mediate PFC activity should occur.

The indication that dopaminergic midbrain activity might play a pivotal role in updating working memory representations hints at a significant overlap in the brain mechanisms responsible for updating representations in working memory and updating learned reward-based values. The same neural circuits have been associated with updating of value-based decision-making and reinforcement learning (Corlett et al., 2022; Jocham et al., 2011; O'Reilly, 2006), in which the dopaminergic midbrain signal known as the reward prediction error (RPE) plays a key role (Montague et al., 1996; Schultz, 2013; Schultz et al., 1997). The RPE signal represents a value-updating signal that occurs when a received reward differs from what was expected according to reinforcement learning models (Miletić et al., 2021; O'Doherty et al., 2017; Sutton & Barto, 2018). The signal is encoded by phasic dopaminergic signals in the VTA and SNc (e.g., Cooper et al., 2012; Montague et al., 1996; O'Doherty et al., 2017; Schultz et al., 1997), targeting PFC representations via the striatum, following the same neural route that has been proposed to be involved in working memory updating (Cohen et al., 2002; Hazy et al., 2007; Murty et al., 2011; Ponzi, 2008).

In this study, we aimed to extend the work of Nir-Cohen et al. (2020) by exploring the neural correlates of working memory updating subprocesses, particularly focusing on the subcortex. To achieve this, we employed a scanning and analysis protocol to improve the signal quality from subcortical nuclei. This protocol included ultra-high field 7 T fMRI with a scanning protocol tailored to meet the requirements for imaging of the subcortex (de Hollander et al., 2017; Miletić et al., 2020) and individually parcellated masks of several subcortical nuclei (Bazin et al., 2020). While Nir-Cohen et al.'s results on striatal gate opening align with the hypothesized role of the striatum in working memory gating , they do not offer enough evidence to definitively rule out the subcortex's involvement in other working memory update subprocesses. Hence, the present study specifically aimed to shed light on subcortical involvement in working memory subprocesses–associated with gating, substitution, and being in an updating mode–and to discern contributions from several subcortical structures. In addition to investigating subcortical nuclei associated with the basal ganglia-thalamo-cortical loop, we hypothesized that midbrain nuclei containing dopaminergic neurons (VTA and SN) might play a crucial role in several working memory updating subprocesses, with their activation in each contrast indicating different neural mechanisms. Precisely, in light of the findings from Nir-Cohen et al. discussed earlier, we tested whether the VTA demonstrates enhanced activity during gate closing as postulated by the dual-state theory (Durstewitz & Seamans, 2008). Furthermore, we aimed to investigate evidence of neural correlates of working memory substitution in the midbrain, specifically in the VTA and/or SN, which would support the idea that a phasic dopaminergic signal contributes to the

act of changing working memory content, akin to the process of value updating following RPE. In contrast, VTA and/or SN activity in the gate opening process would lend support to theoretical accounts suggesting that a phasic dopaminergic signal opens the gate to working memory and, this way, facilitates (potential) subsequent working memory updating. Along these lines, activation in the VTA and/or SN during an updating mode may support the idea that dopaminergic signals are actively engaged in keeping the gates to working memory open.

Taken together, this 7T fMRI study was designed to shed light on subcortical–in particular dopaminergic–contributions to working memory updating subprocesses in the human brain.

# 2 Methods

## 2.1 Procedure

37 participants (20 female; mean age 26.65± 5.72 years; age range 19 – 39 years) took part in the study, which was approved by the ethical committee at the University of Amsterdam, the Netherlands and the Regional Committees for Medical and Health Research Ethics, Norway. All participants provided written informed consent and completed MRI screening forms to ensure they were eligible for scanning. The recruitment was conducted at the Norwegian University of Science and Technology. The participants had a corrected-to-normal vision and no history of epilepsy or overt clinical neuropsychiatric disease. Two participants' data were excluded because they had more than 30 percent non-responses or less than 70 percent accuracy on the reference-back task, suggesting that these participants were either disengaged or misunderstood the task instructions.

## 2.2 Reference-back task

To disentangle the various working memory updating subprocesses, participants completed the reference-back task (Rac-Lubashevsky & Kessler, 2016a). The task required participants to compare a presented stimulus–a capital letter, akin to Rac-Lubashevsky and Kessler, 2016a, 2016b–to a reference stimulus held in working memory (i.e., the referent; Fig. 1). The color of the stimulus frame indicated whether to open the gate to working memory facilitating potential updating of the referent with the presented stimulus (on red frame/reference trials) or to keep the gate closed to maintain the existing referent (on blue frame/comparison trials). When the presented stimulus matched the referent, participants were instructed to respond "same" (by pressing the right key). When the presented stimulus did not match the referent, participants were instructed to respond "different" (by pressing the left key). In other words, reference trials required updating working memory because the current stimulus served as a referent for subsequent trials. Here, the actual act of updating, i.e., the process of replacing old information with new, is represented by substitution. By contrast, comparison trials did not require updating working memory. Working memory gate opening was needed in reference trials that followed comparison trials, and working memory gate closing was required in comparison trials that followed reference trials; hence, both involved a switch in the gate state. Accordingly, the gate state 'switched' or 'repeated' in each trial. This resulted in eight conditions: 2 (trial type: reference vs. comparison) × 2 (switch type: repeat vs. switch) × 2 (response: same vs. different), which formed the basis of the various reference-back contrasts.

In each block of the task, the trial sequence began with a reference trial that required no response. Following the reference trial, a fixation cross was displayed at the center of the display, followed by a framed letter ("X" or "O") presented for a duration of two seconds. Participants were instructed to respond with either "same" or "different" during the stimulus presentation phase. Between each pair of stimuli, a fixation cross was presented. The duration of the fixation cross varied randomly and was selected from a pseudo-exponential distribution, with possible durations of 0.75, 1.5, 2.5, or 3 seconds, in order to decorrelate the design matrix. After the stimulus presentation phase, an intertrial interval was introduced, also in the form of a fixation cross. This intertrial interval had a variable duration ranging from 0.52 to 2.77 seconds in order to ensure that each trial's total duration was set at four times the TR (Repetition Time), totaling 5.53 seconds per trial. This design resulted in a range of total fixation cross durations between 1.27 and 5.77 seconds for each trial. Within each block, all eight conditions formed by the combination of trial type, switch type, and response were presented 16 times, leading to a total of 256 trials (32 trials for each condition).

In accordance with Nir-Cohen et al. (2020), we defined four contrasts based on the eight reference-back conditions (Fig. 2): gate opening, gate closing, substitution, and updating mode. Gate opening was measured by the difference between reference/switch and reference/repeat trials. This is because the process of opening the gate to working memory should only take place in reference/switch trials, as no change in the gate state is required in repeated reference trials. Similarly, gate closing was measured by the difference between comparison/switch and comparison/repeat trials. Substitution takes place in reference/different trials. However, in order to isolate it from any gate-switching effects, the contrast only takes into account 'repeat' trials. Furthermore, in order to set apart the general effects of "same" and "different" responses, the difference between "different" and "same" responses in comparison trials was used as a baseline. Thus, the cost of substitution is indicated by a larger difference between "different" and same responses in reference/repeat trials compared to comparison/repeat trials. Additionally, we computed the updating mode contrast (Nir-Cohen et al., 2020), operationalized as the difference between reference/repeat and comparison/repeat trials (Fig. 2).

## 2.3 Behavioral analysis

Statistical tests on all the conditions, including four a priori contrasts of interest, were conducted to examine the effects of different trial types and reference-back measures on mean response time (RT) and accuracy.

The first trial from each block, which did not require a response, was excluded, as well as any trials with response times faster than 0.150 seconds or slower than 3 seconds, following the exclusion criteria used in Boag et al. (2021). For RT analysis, error trials were excluded.

Linear mixed models were employed to assess the statistical significance of trial type (reference/comparison), switch type (switch/repeat), and response (same/different), along with the four contrasts (gate opening, gate closing, substitution, and updating mode) on mean RT and accuracy. General linear mixed models with a Gaussian link function were used for mean RT, and generalized linear mixed models with a probit link function were used for accuracy. Each model included trial type, switch type, and response as fixed effects, along with random intercepts for each participant. We used a significance criterion of alpha equivalent to .05.

## 2.4 MRI data acquisition

Each participant was scanned four times using a 7 T Siemens MAGNETOM TERRA scanner (gradient strength = 80 mT/m at 200 T/m/s) equipped with a 32-channel Nova Medical's single channel transmit 32-channel receive head coil. Anatomical scans were collected in the first session, and the remaining sessions involved different functional scans. The data discussed in this article were collected in one of the functional sessions that included four functional runs in which the participants completed two different tasks. One task was the reference-back (Fig. 1), which consisted of two runs of 129 trials each. The other task was a reversal-learning task and is not analyzed in the present article. The order of the two tasks was randomized between subjects.

The anatomical session involved acquiring a multi-echo gradient-recalled echo scan (GRE) and an MP2RAGE scan. The GRE scan parameters were as follows: TR = 31.0 ms, TE1 = 2.51 ms, TE2 = 7.22 ms, TE3 = 14.44 ms, TE4 = 23.23 ms, FA = 12°, and FOV = 240 × 240 × 168 mm. The MP2RAGE scan parameters were TR = 4300 ms, TE = 1.99 ms, inversions TI1 = 840 ms, TI2 = 3270 ms, flip angle 1 = 5°, flip angle 2 = 6°, FOV = 240 x 240 x 168 mm, and bandwidth (BW) = 250 Hz/Px (Marques et al., 2010).

The experimental session comprised four functional echo-planar imaging (EPI) runs with four EPI volumes acquired with opposite phase encoding directions for susceptibility distortion correction. The functional data were collected using a single-echo 2D-EPI BOLD sequence with the following parameters: TR = 1380 ms, TE = 14 ms, MB = 2, GRAPPA = 3, voxel size = 1.5 mm isotropic, partial Fourier = 6/8, flip angle = 60°, MS mode = interleaved, FOV = 192 x 192 x 128 mm, matrix size = 128 x 128, BW = 1446 Hz/Px, slices = 82, phase encoding direction = A >> P, and echo spacing = 0.8 ms. Each of the two tasks of the functional session had a total of two runs, with each run lasting 13 minutes and 45 seconds, resulting in a total of four runs and 55 minutes. To help co-register the functional scans to the high-resolution data from the anatomical session, another anatomical MP2RAGE scan (1mm) was collected at the end of the functional session.

Physiological data, including heart rate and respiration, were recorded for all participants to assess the impact of physiological noise on the fMRI data.

## 2.5 MRI data preprocessing

The imaging data was preprocessed using the neuroimaging preprocessing tool *fMRIPrep 20.2.0* (Esteban et al., 2018; Esteban et al., 2019), which is based on *Nipype 1.7.0* (Gorgolewski et al., 2011, 2018; RRID:SCR_002502). The anatomical data preprocessing involved multiple steps such as intensity and non-uniformity correction (using *N4BiasFieldCorrection*, Tustison et al., 2010), skull-stripping (using Nipype's *antsBrainExtraction.sh*), and tissue segmentation (using *FSL*'s fast, Zhang et al., 2001) of the T1-weighted images. The brain-extracted T1-weighted scans were normalized by means of volume-based spatial nonlinear registration to standard space *ICBM 152 Nonlinear Asymmetricl template version 2009c (MNI152NLin2009cAsym*; Fonov et al., 2009, RRID:SCR_008796) using *antsRegistration* (*ANTs 2.3.3*). For more information on anatomical data preprocessing, see Miletić (2023).

The following preprocessing was performed for each of the two functional (BOLD) runs per task per participant. A reference volume and its skull-stripped version were generated by aligning and averaging 1 single-band reference (SBRefs). A B0-nonuniformity map (or fieldmap)

was estimated based on two EPI references with opposing phase-encoding directions, with *3dQwarp* (Cox & Hyde, 1997; AFNI 20160207). Based on the estimated susceptibility distortion, a corrected EPI reference was calculated for a more accurate co-registration with the anatomical reference. The BOLD reference was then co-registered to the T1w reference using *bbregister* (FreeSurfer 6.0.1), which implements boundary-based registration (Greve & Fischl, 2009). Co-registration was configured with six degrees of freedom. Head-motion parameters with respect to the BOLD reference (transformation matrices and six corresponding rotation and translation parameters) were estimated before any spatiotemporal filtering using *mcflirt* (FSL 5.0.9, Jenkinson et al., 2002). BOLD runs were slice-time corrected using *3dTshift* from AFNI 20160207 (Cox & Hyde, 1997; RRID:SCR_005927). A reference volume and its skull-stripped version were generated using a custom methodology of *fMRIPrep*. The BOLD time-series (including slice-timing correction when applied) were resampled onto their original, native space by applying a single, composite transform to correct for head-motion and susceptibility distortions. These resampled BOLD time-series will be referred to as preprocessed BOLD in original space, or just preprocessed BOLD. Several confounding time-series were calculated based on the preprocessed BOLD: framewise displacement (FD), 'DVARS' (the spatial standard deviation of difference images), and three region-wise global signals. FD was computed using two formulations following Power (absolute sum of relative motions, Power et al., 2014) and Jenkinson (relative root mean square displacement between affines, Jenkinson et al., 2002). FD and DVARS were calculated for each functional run, both using their implementations in *Nipype* (following the definitions by Power et al., 2014). The three global signals were extracted within the CSF, the WM, and the whole-brain masks. Additionally, a set of physiological regressors was extracted to allow for component-based noise correction (*CompCor*, Behzadi et al., 2007). Principal components were estimated after high-pass filtering the preprocessed BOLD time-series (using a discrete cosine filter with 128 seconds cut-off) for the two *CompCor* variants: temporal (*tCompCor*) and anatomical (*aCompCor*). *tCompCor* components were then calculated from the top 2 percent variable voxels within the brain mask. For *aCompCor*, three probabilistic masks (CSF, WM, and combined CSF + WM) were generated in anatomical space. The implementation differs from that of Behzadi et al. (2007). Instead of eroding the masks by two pixels on BOLD space, the *aCompCor* masks were subtracted from a mask of pixels that likely contain a volume fraction of GM. This mask is obtained by dilating a GM mask extracted from the FreeSurfer's *aseg* segmentation, and it ensures that components were not extracted from voxels containing a minimal fraction of GM. Finally, these masks were resampled into BOLD space and binarized by thresholding at 0.99 (as in the original implementation). Components were also calculated separately within the WM and CSF masks. For each *CompCor* decomposition, the k components with the largest singular values were retained, such that the retained components' time-series are sufficient to explain 50 percent of variance across the nuisance mask (CSF, WM, combined, or temporal). The remaining components were excluded. The head-motion estimates calculated in the correction step were also placed within the corresponding confounds file. The confound time-series derived from head motion estimates and global signals were expanded with the inclusion of temporal derivatives and quadratic terms for each (Satterthwaite et al., 2013). Frames that exceeded a threshold of 0.5 mm FD or 1.5 standardized DVARS were annotated as motion outliers. All resamplings can be performed with a single interpolation step by composing all the pertinent transformations (i.e., head-motion transform matrices, susceptibility distortion correction when available, and co-registrations to anatomical and output spaces). Gridded

(volumetric) resamplings were performed using *antsApplyTransforms* (ANTs), configured with Lanczos interpolation to minimize the smoothing effects of other kernels (Lanczos, 1964). Non-gridded (surface) resamplings were performed using *mri_vol2surf* (FreeSurfer). Many internal operations of fMRIPrep use *Nilearn 0.6.2* (Abraham et al., 2014, RRID:SCR_001362), mostly within the functional processing workflow.

## 2.6 Regions-of-interest

For the selection of regions-of-interests (ROIs), we selected only subcortical masks for our ROI analyses given our main research aim centered on the subcortex. Second, we included individual subcortical masks derived from the MASSP automated parcellation algorithm (Bazin et al., 2020) to increase delineation accuracy on a subject level. Third, bilateral masks for each ROI were included to maximize regional specificity. Fourth, the pallidum was included with individual masks for its external (GPe) and internal (GPi) segments. Fifth, playing an important role in subcortical functioning in general and in basal ganglia processes in particular, the subthalamic nucleus (STN) mask was also selected for this study. Most importantly, we incorporated masks of the dopaminergic midbrain given our secondary research aim focused on the dopamine sources. This resulted in 7 masks (14 bilaterally). Hence, the MASSP algorithm (Bazin et al., 2020) was used to parcellate individual anatomical masks for the thalamus (Tha), striatum (Str), GPe and GPi, STN, SN, and VTA.

In addition, for a post-hoc ROI analysis we extended the dopaminergic midbrain masks utilizing the subdivision masks based on the probabilistic atlas from Pauli, Nili, and Tyszka (2018). These masks diverge from the delineations by MASSP (Bazin et al., 2020) in both shape and volume (Trutti et al., 2021), and were constructed with the aim of delineating two functionally different component nuclei of the VTA. Specifically, the nucleus of the VTA (VTAnc) and Parabrachial Pigmented (PBP) nucleus. The VTAnc refers to a small ventromedially located nucleus of the VTA, and the PBP mask represents a dorsolateral component nucleus of the VTA. Both VTAnc and PBP masks provided by Pauli et al. (2018) fall into the region defined as VTA by MASSP (Trutti et al., 2019). In addition, the authors put forward a subdivided SN, providing separate masks of the SNc and SNr, respectively (Pauli et al., 2018). Given that both VTA masks and the SNc are associated with dopaminergic cell populations, ROI analyses were repeated employing these three masks.

Furthermore, a supplementary exploration into the BOLD signal clusters within ROIs was carried out for comparison with Nir-Cohen et al. (2020). This analysis utilized the whole-brain statistical parametric maps (SPMs) derived from the whole-brain GLMs and, therefore, did not include individual parcellations but, aligning with Nir-Cohen et al. (2020), employed probabilistic masks of the striatal subdivision putamen (Pu) and caudate nucleus (Ca) from Pauli et al. (2018), instead of the entire striatum from MASSP. The remaining masks were taken from the probabilistic group-level masks from Bazin et al. (2020). See the appendix for the cluster-based ROI analysis elaboration (appendix A1.1).

All masks were registered to *MNI152NLin2009cAsym* using *antsRegistration* (*ANTs 2.3.3*).

## 2.7 fMRI statistical analysis

The aim of our 7 T fMRI study was to shed light on subcortical–in particular possible dopaminergic–contributions to working memory updating subprocesses in the human brain. To achieve this, we extended Nir-Cohen and colleagues' (2020) fMRI data analysis procedure, with a specific emphasis on the subcortical regions. Our foundational objective was to compare the results obtained from a protocol optimized for BOLD sensitivity in the subcortex with the current empirical evidence, which has not definitively established the involvement of subcortical regions. This included a multi-step fMRI analysis. First, a whole-brain analysis was conducted to investigate brain activation by means of SPMs for each contrast on a whole-brain level. Second, ROI-wise GLM analysis with increased regional specificity due to extraction from unsmoothed data was conducted to explore the contribution of several subcortical nuclei in working memory updating. Details about the GLM analysis on both the whole-brain and region-specific levels are listed below. Third, a cluster-based ROI analysis was carried out to explore the presence of significant clusters within each ROI, akin to Nir-Cohen et al. (2020) (for detailed methodology, see A1.2 in the appendix).

A canonical double gamma hemodynamic response function (HRF) with temporal derivative was employed as the basis set for all methods of analysis (Glover, 1999). The design matrix was constructed to encompass the eight experimental conditions associated with the reference-back resulting from the 2 (trial type: reference vs. comparison) × 2 (switch type: repeat vs. switch) × 2 (response type: same vs. different) factorial design. Error trials were not included in the fMRI analysis. Before conducting the GLM analysis, the functional data underwent high-pass filtering (Smith & Brady, 1997), and subsequently, spatial smoothing was applied using SUSAN (kernel-size full-width half-maximum = 4.5 mm). Notably, this spatial smoothing factor deviates from the one employed in the study by Nir-Cohen et al. in 2020, where a larger smoothing factor of 6 mm was used. This variation arises due to differences in image acquisition protocols and the improved image resolution and quality in our study, making such extensive smoothing unsuitable for our dataset (De Hollander et al., 2015). In addition to the task-specific regressors, our design matrix incorporated six motion parameters (comprising three translational and three rotational parameters), along with DVARS and FD estimates obtained during preprocessing. An 18-regressor RETROICOR model (Glover et al., 2000) was employed to model physiological noise. The model comprised a fourth-order phase Fourier expansion of the heart rate signal, a second-order phase expansion of the respiration signal, and a second-order phase Fourier expansion of the interaction between heart rate and respiration (Harvey et al., 2008). Additional regressors were used to account for heart rate variability (HRV; Chang et al., 2009) and respiratory volume per time unit (RVT; Birn et al., 2008; Harrison et al., 2021). The PhysIO toolbox (Kasper et al., 2017) implemented in the TAPAS software (Frässle et al., 2021) was employed to estimate the physiological regressors. For two participants, either both runs or only the second run of physiological data were not collected due to technical issues. In these cases, we substituted the first 20 components obtained through aCompCor, as described by Behzadi et al. (2007). Therefore, a total of 36 regressors were used in the model.

Whole-brain analyses were conducted using the FILM method from FSL FEAT (Jenkinson et al., 2012; Woolrich et al., 2001). These analyses took into account autocorrelated residuals. For the purpose of combining the GLMs at the run level for each task, fixed effects analyses were employed. Group-level models were subsequently estimated using FLAME1+2

from FSL (Woolrich et al., 2001). SPMs were generated to visualize the resulting group-level models. The maps were corrected for the family-wise error rate (FWER) using the random Gaussian field (GRF) procedure and a critical value of $q < 0.05$ (Nichols & Hayasaka, 2003), and a minimum cluster size ($K_E$) of 10 voxels.

In addition, the ROI GLM analysis was conducted as follows: Mean time-series data were extracted from each subcortical region of interest first, using probabilistic masks provided by the individual MASSP parcellation (Bazin et al., 2020) and second, using the probabilistic (non-individual) atlas by Pauli et al. (2018). Each voxel's contribution to the mean signal of the region was weighted based on its probability of belonging to that region. Subsequently, the time-series data were transformed into percentage signal change values by dividing each timepoint by the mean signal of the time-series, multiplying the result by 100, and then subtracting 100. These time-series data were extracted from unsmoothed data to ensure regional specificity. The individual runs were concatenated for analysis. We infer exclusively from positive BOLD responses, given the disagreements surrounding negative BOLD responses (Schridde et al., 2008; Wade, 2002). The ROI GLMs thus encompassed ROI-wise GLM's fit on the mean time-series extracted from the unsmoothed functional data of each voxel. Bayesian one-sample *t*-tests were computed to investigate the evidence for each contrast based on the mean beta (i.e., signal change) derived from the ROI GLMs. For classification, we utilized Bayesian classification by Jeffreys (1998), where Bayes Factors between 1 and 3 indicate weak evidence, Bayes Factors between 3 and 10 indicate moderate evidence, and Bayes Factors greater than 10 indicate strong evidence.

Lastly, by means of providing comparability with the only other existing fMRI study employing the same experimental paradigm (Nir-Cohen et al., 2020), we also ran an additional 'cluster-based ROI analysis' which examined each ROI for clusters of activation based on the whole-brain GLMs after correcting for multiple comparisons across the ROI's voxels. Here, we investigated the involvement of specific brain regions previously associated with working memory, including the frontoparietal network and the basal ganglia in each working memory subprocess (for methods and results, see A1 in the appendix).

# 3 Results

## 3.1 Behavioral Results

The results of the behavioral analyses of the reference-back task are presented in the appendix (Tab. A3 – A5). Descriptive statistics are reported in Table 1.

Figure 2 illustrates the group-averaged mean RT and accuracy for each condition of the experimental design. The overall mean RT was 0.86 seconds (mean RT for correct responses was 0.85 seconds), and the overall accuracy on the task was 94.7 percent, indicating that participants understood the task instructions. Correct responses were, on average, faster than errors ($\Delta$ = 0.169 s, $t = 12.8$, $df = 149$, $p < .001$, $d = 0.61$).

For mean RT, there were significant main effects of trial type, switch type, and response. Additionally, the two-way interactions but not the three-way interactions were significant. For

accuracy, there were no significant main effects. However, we found two-way interactions between trial type and switch type paired with trial type and response. There was no three-way interaction (Tab. A3-A5, Fig. A2).

Responses were slower and less accurate on reference/switch (RT: $M$ = 0.916 s, $SEM$ = 0.017; Accuracy: $M$ = 0.920, $SEM$ = 0.008) than reference/repeat trials (RT: $M$ = 0.856 s, $SEM$ = 0.016; Accuracy: $M$ = 0.962, $SEM$ = 0.005), which represents the behavioral costs of working memory gate opening.

Responses were not substantially slower on comparison/switch ($M$ = 0.829 s, $SEM$ = 0.012) than comparison/repeat trials ($M$ = 0.821 s, $SEM$ = 0.008), suggesting no behavioral RT cost of closing the working memory gate. Interestingly, responses were more accurate on comparison/switch trials ($M$ = 0.962, $SEM$ = 0.005) compared to comparison/repeat trials ($M$ = 0.933, $SEM$ = 0.008), indicating that gate closing did not produce any behavioral cost, in fact an accuracy gain in keeping the gate closed.

For the substitution contrast, the difference between "different" and "same" responses on repeated reference trials was slower and less accurate (RT: $\Delta$ = 0.207 s, $SEM$ = 0.017; Accuracy: $\Delta$ = -0.019, $SEM$ = 0.010) than the difference between "different" and "same" responses in repeated comparison trials (RT: $\Delta$ = 0.082 s; $SEM$ = 0.010; Accuracy: $\Delta$ = 0.007, $SEM$ = 0.012). This suggests costs of both response time and accuracy in working memory substitution processing.

Finally, responses were slower but more accurate on repeated reference trials (RT: $M$ = 0.856 s, $SEM$ = 0.016; Accuracy: $M$ = 0.962, $SEM$ = 0.005) responses compared with repeated comparison trials (RT: $M$ = 0.821 s, $SEM$ = 0.008; Accuracy: $M$ = 0.933, $SEM$ = 0.008), reflecting the response time cost of being in a general updating mode.

## 3.2 fMRI Results

### 3.2.1 Whole-brain GLMs

Whole-brain analyses were conducted to explore the brain activations associated with gate opening, gate closing, substitution, and updating.

#### Gate Opening

Gate opening was associated with large bilateral clusters of activation across the cortex. Specifically, we found increased activation in frontal cortical regions, including the mPFC and dlPFC, as well as the somatosensory and motor cortices (Fig. 4; Tab. 2). Gate opening was also associated with increased activation in posterior parietal regions, including the precuneus cortex and left parietal lobe. The occipital lobe, including the left and right fusiform cortex, also showed increased activation during gate opening. Subcortically, gate opening was associated with clusters of increased activation in the thalamus, in particular a large cluster in the left thalamus. These findings are consistent with Nir-Cohen et al. (2020). Additionally, we found a large cluster of activation in the brainstem and midbrain regions (see Fig. 8), covering areas such as the left and right red nucleus, periaqueductal gray, superior peduncle, and left subthalamic nucleus.

### Gate Closing

Gate closing was associated with a few clusters of cortical activations, in line with Nir-Cohen et al. (2020). Our largest cluster covered much of the left PPC; another cluster occurred in the left dlPFC. In contrast to findings by Nir-Cohen et al., there was no activation in the right hemisphere except for relatively small activation clusters in the bilateral occipital fusiform gyrus. In addition, gate closing was associated with increased activity in the left inferior temporal gyrus and cerebellum (Fig. 5; Tab. 2).

### Substitution

Substitution was associated with several large clusters of activation. There was increased activation in the premotor cortex and a substantial portion of the PFC, including BA8 (dlPFC), BA24, and BA32 (mPFC). There was increased activation in the pre-supplementary motor area (preSMA), the superior and inferior frontal gyrus, the left and right PPC, and the left inferior parietal lobule (IPL). The subcortex showed heightened activation in both the right and left striatal regions, including the caudate and putamen. A complete list of all active clusters is given in Table 2. The pattern of activation associated with substitution is broadly consistent with Nir-Cohen et al. (2020). However, we found a considerably greater extent of cortical activation, along with subcortical activation not observed by Nir-Cohen et al. (2020) (Fig. 6 and 8).

### Updating Mode

The updating mode was associated with increased activation in frontal and posterior parietal regions. The frontal activation covered the left and right medial to dorsolateral prefrontal regions (from the preSMA, over the superior and middle frontal gyrus, to the inferior frontal gyrus). The posterior parietal cluster covered the IPL and stretched from the supramarginal gyrus, medially toward the precuneus cortex, and ventrally along the angular gyrus toward the medial temporal gyrus. There was increased activation in occipital regions, including the fusiform gyrus and the intra- and supracalcarine cortex. In the subcortex, clusters of increased activation were found in the left and right putamen, caudate nuclei, and the right thalamus. This pattern of activation is consistent with Nir-Cohen et al. (2020) but extends their findings to a broader range of cortical (Fig. 7; Tab. 2) and subcortical regions (Fig. 8).

### 3.2.2 ROI Analyses

We conducted three sets of ROI analyses, as described in 2.6. The main ROI analyses included ROI-wise GLMs based on the individually parcellated masks derived from MASSP (Bazin et al., 2020). In addition, we conducted an additional ROI analysis into subdivisions of the dopaminergic midbrain regions through the exploration of masks derived from a probabilistic atlas. A supplementary exploration into clusters of activation within ROIs based on the whole-brain GLMs was also performed for comparison with Nir-Cohen et al. (2020) (see A1.1) Hence, the first ROI-wise GLMs ('ROI-wise GLMs on individually parcellated masks') emphasized individual delineations and thus provided high precision with regard to individual anatomy. While the post-hoc ROI-wise GLMs provided information on the possible differential involvement of subnuclei of the VTA and SN, however, at the cost of individual anatomical detail and will be referred to as 'ROI-wise GLMs based on VTA/SN subdivisions atlas' in the following. Results of the ROI analyses are illustrated in Figures 9 and 10, Tables 3 and 4.

#### 3.2.2.1 ROI-wise GLMs on individually parcellated masks

ROI-wise GLMs analyses demonstrated that subcortical regions play a role in each experimental contrast. The main findings for each experimental contrast are summarized below and depicted in Figure 9 and Table 3.

**Gate opening**

During gate opening, subcortical activation was limited to the right thalamus. Bayesian analyses provided only weak evidence for activity in the right thalamus, contrasting the observation of extensive bilateral thalamus activation in the whole-brain contrast. Notably, we found moderate evidence against activity in any other ROI in the basal ganglia and midbrain, along with weak evidence against activity in the right thalamus, right GPe, and right GPi.

**Gate closing**

ROI-wise GLM results indicated moderate evidence for activity in the right GPe and weak evidence for activity in the right basal ganglia nuclei, namely right striatum and right STN, during gate closing. Moreover, bilateral midbrain activation of the SN was observed, indicated by weak evidence.

**Substitution**

During substitution, the results of ROI-wise GLMs confirmed the engagement of subcortical structures, although there were variations in their individual contributions. Moderate evidence of activity was observed in the striatum bilaterally, in the right STN, and the right VTA. Moreover, weak evidence indicated activity in the thalamus, left GPe, left GPi, and left STN. Overall, Bayesian analyses did not indicate strong evidence against activity in any subcortical ROI during the process of substitution.

**Updating mode**

Being in an updating mode was associated with both basal ganglia and midbrain activity. Precisely, ROI-wise GLM results provided Bayes factors suggesting strong evidence for activity in the left GPe and right STN. Yet, only moderate evidence for the right GPe, left GPi, and right SN. Furthermore, weak evidence from the ROI GLMs indicated activity of the thalamus, striatum, left STN, and right VTA. Overall, there were no Bayes factors indicating strong evidence against activity related to the updating mode in any ROI.

#### 3.2.2.2 ROI-wise GLMs based on VTA/SN subdivisions atlas

ROI-wise GLM analysis demonstrated that midbrain nuclei play a role in gate closing, substitution, and updating mode. The main findings for each experimental contrast are summarized below and depicted in Figure 10 and Table 4.

**Gate opening**

Confirming results from the preceding ROI-wise GLM analysis, ROI-wise GLMs using the probabilistic atlas masks implicated no evidence for activity in midbrain dopamine-producing nuclei in gate opening, supported by moderate evidence against activity in any mask from Pauli et al. (2018).

**Gate closing**

ROI-wise GLMs provided evidence for SNc activation during gate closing. There was strong evidence for activity in the right SNc and moderate evidence for activity in the left SNc. This high Bayes factor in favor of the right SNc suggests a dopaminergic involvement during gate closing. Furthermore, the results of the Bayesian analyses did not provide strong evidence against activity in any ROI during gate closing trials.

## Substitution

In line with the results from the ROI analyses based on MASSP delineations (see above), results implicated activity in the right VTA. Specifically, the right PBP mask by Pauli et al. (2018) indicated activity with moderate evidence, suggesting an involvement of the dorsolateral nucleus of the VTA during substitution. In addition, BOLD signal change in the right SNc demonstrated a high Bayes factor. Both findings indicate dopaminergic involvement in the substitution process.

## Updating mode

Midbrain masks by Pauli et al. (2018) demonstrated an extremely high Bayes factor for the right SNc–insinuating very strong evidence for increased activity– moderate evidence for activation in the left SNc, along with weak evidence for activity in the right PBP during updating mode. Again, there were no Bayes factors that indicated strong evidence against activity related to the updating mode in any ROI.

In summary, no subcortical engagement except for the thalamus was revealed in the gate opening contrast. However, ROI-wise GLMs indicated only weak evidence for the right thalamus during gate opening, rendering our results on gate-opening-related thalamic activity inconclusive. Notably, we observed moderate evidence against the involvement of most ROIs in the basal ganglia and midbrain during gate opening. Nevertheless, during gate closing, substitution, and updating mode, evidence did suggest activity in the basal ganglia and midbrain. It appears, however, that basal ganglia nuclei are differently engaged across gate closing, substitution, and updating mode contrasts. Reasonable evidence for striatal engagement was only found during working memory substitution trials. In the pallidum, gate closing-associated activity implicated the right GPe, and there was very strong evidence for the left GPe in updating mode. Additionally, results implicated the involvement of the right STN in updating mode and in substitution.

Also, the regions of the midbrain appear to be differently engaged across the three contrasts. The right SN was active in the updating mode while there was only weak evidence for increased SN activation during gate closing and none during substitution. However, using the probabilistic mask by Pauli et al. (2018), data indicated strong evidence for SNc activation during updating mode and strong evidence for SNc engagement during gate closing and substitution. Interestingly, evidence for engagement of the left SNc was substantially lower in all three contrasts. Furthermore, results implicated the right VTA mask by MASSP (Bazin et al., 2020) in substitution. Interestingly, considering the analyses using the probabilistic masks, this activation appears to be driven by the (right) PBP, a VTA component nucleus.

# 4 Discussion

The present study aimed to shed light on the neural substrates of working memory subprocesses, particularly focusing on the subcortex. As an extension of the work by Nir-Cohen et al. (2020), we employed the reference-back paradigm in conjunction with 7 T fMRI, including a scanning and analysis protocol optimized for the subcortex, to precisely discern contributions from several subcortical structures to working memory updating subprocesses associated with gating, substitution, and being in an updating mode. In addition to investigating nuclei in the basal ganglia-thalamo-cortical loop, we hypothesized that midbrain nuclei containing dopaminergic neurons, the VTA, and SN might play a pivotal role in working memory updating subprocesses.

The whole-brain analysis not only revealed a substantially broader range of brain activation in both the cortex and subcortex but also provided more detail compared to previous work (Nir-Cohen et al., 2020). More precisely, consistent with the work by Nir-Cohen et al. (2020), we found that substitution and being in a general mode of updating show increased activation in all regions belonging to the frontoparietal network (FPN). Overall, the components of the FPN seem to be involved across all contrasts. However, our data suggests that each working memory subprocess differently engages the individual FPN components: posterior parietal regions play a greater role in gate-switching, while substitution and updating primarily recruit the frontal regions and subcortex. This observation implies that gate-switching may primarily serve as a selective attention process, a view supported by Schouwenburg et al. (2014) through their findings on attention-modulated cortical neural activities involved in gate-switching mechanisms.

Furthermore, our whole-brain analysis identified thalamic activation in the process of opening the gate to working memory, working memory substitution, and the updating mode–aligning with Nir-Cohen et al.'s findings. Nonetheless, they found thalamic activity only in a region-of-interest (ROI) analysis and not their whole-brain analysis, demonstrating our enhanced signal and spatial resolution. Additionally, our whole-brain analysis revealed activation in distinct subcortical regions: the brainstem during gate opening, cerebellar regions during gate closing, and midbrain during gate opening, substitution, and updating mode. Notably, during substitution and updating mode, there was activation in striatal subregions–specifically the caudate nucleus and putamen–particularly extensive bilateral striatal activation during substitution, aligning with the PBWM model's actual updating process.

We extend previous work by investigating individual masks of subcortical ROIs using the automated subcortical parcellation algorithm MASSP (Bazin et al., 2020), including regions covering the dopaminergic midbrain–additionally supported by ROI analyses using masks from Pauli et al. (2018)–each considered bilaterally. Together, they revealed activation in the basal ganglia and midbrain during gate closing, substitution, and updating mode, albeit with differing degrees of engagement across these contrasts. Notably, no subcortical engagement was observed in gate opening, except for weak evidence for the right thalamus. This still might indicate a possible involvement of the thalamus during gate opening since the whole-brain analysis yielded large bilateral activation clusters in the thalamus. Furthermore, while the Bayesian analysis

provided moderate evidence against any subcortical ROI in gate opening, the evidence against the right thalamus was nihil. All these factors render the results for thalamic involvement during gate opening inconclusive. However, it should be noted that activation limited to a thalamic subnucleus may explain the observed findings from the ROI analysis. Precisely, it is possible that an existing activation was negated as the BOLD signal change was averaged across all voxels included in a specific mask. Our data, therefore, did not definitively support or contradict thalamic involvement in gate opening, underscoring the necessity for individually delineated masks for subnuclei in the thalamus in future studies.

Our findings refine prominent neural theories of working memory gating by showing that, rather than controlling gating, the basal ganglia's role may be more specific to the actual act of updating working memory representations with new information and a sustained open-gate state. The reasons for this are twofold. First, there was an absence of evidence for striatal participation in gate opening. Second, we found midbrain and basal ganglia activation during substitution, suggesting a neural signature of midbrain activation–potentially dopaminergic–of the basal ganglia–thalamo-cortical loop. These findings significantly advance our understanding of the cortical and subcortical neural basis of working memory updating subprocesses.

## No support for striatal gate opening

The first aspect of the twofold revelation, indicating a role for the basal ganglia in working memory updating that differs from what was originally postulated, is the absence of evidence for striatal gate opening. We observed no increased activation in the striatum during gate opening but primarily activations in FPN regions, suggesting no active striatal involvement in opening the gate to working memory. In fact, ROI-wise GLMs suggest evidence against the involvement of any basal ganglia nucleus in gate opening. This contrasts with the findings of Nir-Cohen et al. (2020) and raises questions about the relationship between the gate opening process in the reference back task and the indirect striatal gating mechanism described in the PBWM model (Frank et al., 2001; Hazy et al., 2007; O'Reilly & Frank, 2006) and other neurocomputational theories (Hazy et al., 2007; Jongkees, 2020). According to these models, a dopaminergic signal in the striatum is required to trigger gating. Although the orthogonal contrasts in the reference-back task are intended to isolate working memory subprocesses inspired by models of working memory, the two gating contrasts do not fully capture the gating mechanism as originally proposed in neurocomputational models (Frank et al., 2001; Hazy et al., 2007; O'Reilly & Frank, 2006).

However, our finding is partially consistent with the dual-state theory's proposal that dopamine directly modulates prefrontal cortex (PFC) representations via mesocortical pathways (Durstewitz & Seamans, 2008) without involving the basal ganglia. Aligning with the dual-state theory (Durstewitz & Seamans, 2008), the whole-brain analysis did suggest the involvement of the dopaminergic midbrain in gate opening (Fig. 8), yet ROI-wise GLM as well as cluster-based ROI analyses could not further corroborate these findings (see appendix). In fact, ROI-wise GLMs indicated moderate evidence against midbrain activity during the process of opening the gate. All analyses thus point to no basal ganglia involvement specific to working memory gate opening. Instead, there is a possibility that mesocortical dopaminergic pathways could regulate gating. However, the evidence for the midbrain involvement remains mixed, leaving the matter

inconclusive and implying future work should follow up on the exact role of the dopaminergic midbrain in gate opening. Exploring the connectivity of the mesolimbic pathway during gate opening, for example, could provide valuable insights.

Another explanation for the lack of enhanced striatal activity in gate opening challenges the conceptualization of the gating mechanism in the reference-back task, which, as mentioned above, does not accurately map onto the PBWM predictions. The gate opening contrast includes reference trials that follow comparison trials regardless of the stimulus match condition. This means that actually gating new, relevant information into the PFC is only necessary on 50% of gate opening trials (i.e., on reference/switch/mismatch trials and not reference/switch/match trials; not to be confused with substitution contrast, which takes into account only repeated trials). Furthermore, in light of the PBWM model's proposal that the basal ganglia sit in a gate-closed state by default, gate opening should take place on every single reference trial. However, engaging in striatal gating every time a reference cue is encountered (i.e., on every reference trial) would be inefficient if the cue is not predictive of the subsequent updating, as is the case for the reference-back task. As a result, this conceptualization of gate opening would be expected to destabilize working memory representations, potentially reducing the accuracy of working memory-based decisions. In support of the idea that striatal gating in response to each reference cue would be an uneconomic brain process, we did not find lower accuracy in reference trials when compared to comparison trials (see appendix, Table A4), suggesting that gate opening does not occur. Since we, however, did find accuracy costs related to gate opening trials, our findings imply that the costs associated with gate opening might be driven by the trial type/mode switch triggering selective attention rather than reflecting processing costs associated with switching the working memory gate state. That gating costs observed in the reference-back task reflect expenses that are not necessarily associated with mechanisms for switching the gate state is further corroborated by the biology of the basal ganglia, which sit in a default closed-gate (No-Go) state. However, suppose a striatal gating process were to occur in response to every reference trial to facilitate working memory updating. In that case, there should also be striatal/basal ganglia activation in trials where the gate is opened independent of a switch in the gate state. This situation is partially represented in the updating mode, which contrasts repeated reference over repeated comparison trials. If the striatum is only involved when the gate switches from closed to open, no striatal activation should exist in an updating mode.

Intriguingly, we found that the updating mode was associated with strong evidence for basal ganglia engagement, along with strong evidence for the right SNc and right VTA–supported by the right PBP–activity. These findings suggest that repeated reference trials, hence repeated updating cues, engage the basal ganglia–which, in fact, was originally postulated for gate opening by the PBWM model–and that such a sustained updating mode, rather than switching between trial types, causes this activation. Hence, there is a striatal mechanism that is engaged when being prepared for updating is (repeatedly) required. However, this finding is inconsistent with Nir-Cohen et al. (2020), who found no evidence for any subcortical involvement in the updating mode contrast. In fact, they found activation contradictory to our results in both the gate opening and updating mode contrasts, rendering future work on the subcortical engagement in these working memory processes essential to understanding the exact neural mechanisms involved. Alternatively, it is also possible that the basal ganglia involvement in the updating

784 mode reflects attentional processes (Cools et al., 2004; Leber et al., 2008; Schouwenburg et al.,
785 2014) triggered by successive reference trials.

786 Taken together, our data aligns with the PBWM model, emphasizing the central role of
787 the basal ganglia in working memory processes overall. However, we could not find any evidence
788 for striatal involvement during gate opening, in contrast to previous work (Nir-Cohen et al.,
789 2020). Instead, our observations revealed the involvement of FPN regions and accuracy costs
790 during gate opening. These findings suggest that the observed gate opening costs may be better
791 characterized as a selective attention process that does not require striatal selective gating
792 mechanisms.

793 Moreover, despite the lack of striatal involvement during gate opening, our findings do
794 not rule out the possibility that the PBWM model's predictions about striatal gating in working
795 memory are correct, given the misalignment between the gate opening contrast and the PBWM's
796 proposal regarding striatal gating. It remains unclear whether the absence of striatal activation
797 during gate opening trials is specific to low-demand tasks, like the reference-back task, which
798 does not require as much gating compared to high working memory-demand tasks involving
799 preparation for updating. Or whether the gate opening contrast does not sufficiently capture the
800 PBWM proposed gating mechanism. Further investigation is needed to determine whether
801 (dopamine-driven) striatal gating occurs in high-demand working memory tasks, where the
802 gating process plays a more critical role.

803 Additionally, our results suggest a more specific role for the basal ganglia in a ready-to-
804 update mode rather than controlling the gate to working memory, as defined in the reference-
805 back task's gate opening contrast. The contrasting patterns of activation observed in the basal
806 ganglia, particularly the striatum, between the gate opening and updating mode in our dataset
807 compared to the findings of Nir-Cohen et al. (2020) highlight an interesting discrepancy and
808 emphasize the importance of additional research to better understand these variations.

## 809 Dopaminergic involvement in working memory substitution

810 The second aspect of the twofold revelation, indicating a revised role for the basal
811 ganglia in working memory, is the observation of basal ganglia activation in combination with
812 midbrain activation, specifically during substitution trials. Our findings suggest that the striatum–
813 along with the rest of the basal ganglia and the thalamus–is more involved in the actual process
814 of replacing and updating working memory representations ('substitution') than controlling the
815 more general sustained ready-to-update state ('updating mode'). This observation is supported by
816 the enhanced subcortical activity we found in substitution trials evident from whole-brain
817 analyses and ROI-wise GLMs (as well as cluster-based ROI analysis, see appendix). Nir-Cohen et
818 al. (2023) further affirm our observation through their modified version of the reference-back
819 paradigm, which includes task switching, demonstrating that the gate opens only when truly
820 necessary (i.e., when updating of the task that is held in working memory is required) and is
821 associated with increased activity in the basal ganglia, thalamus, and midbrain.

822 We found enhanced neural activation in the dopaminergic midbrain, basal ganglia,
823 thalamus, and PFC when substituting old with new information in working memory. In greater
824 detail, the whole-brain analysis showed large cortical and subcortical activation clusters. Further

825 confirmation from ROI analyses indicated the involvement of the basal ganglia-thalamo-cortical
826 loop in substituting information in working memory. Precisely, ROI-wise GLMs suggested
827 evidence for increased activity in the striatum, right STN, and right VTA, along with activity in
828 the right SNc and right PBP. Furthermore, the cluster-based ROI analysis revealed the
829 involvement of the dlPFC and mPFC, thalamus, and basal ganglia, as well as VTA and SN, both
830 bilaterally notably (see Tab. A5).

831 It is important to note the differences between the ROI-wise GLMs and the cluster-
832 based ROI analysis (see appendix). The cluster-based ROI analysis examines peak activation in a
833 cluster-based approach, just like whole-brain GLMs, but limited to the ROI. Meanwhile, the
834 ROI-wise GLMs consider the mean signal change over all voxels within a mask. Hence, it is
835 possible that the observed signal change found for both left and right midbrain nuclei in cluster-
836 based ROI analysis for substitution was canceled out in the ROI-wise GLMs, resulting in
837 activation observed only in the right VTA, PBP, and SNc during substitution, but in none of the
838 individually delineated SN mask from MASSP (Bazin et al., 2020), which also covers the region
839 associated with the non-dopaminergic reticulata part of the SN (SNr). Future studies can further
840 infer dopamine involvement during substitution by tracking the dopaminergic dynamics in
841 binding and release at receptor sites quantified with positron emission tomography (PET)
842 (Bäckman et al., 2017). Nonetheless, all three analyses imply engagement of the basal ganglia-
843 thalamo-cortical loop, along with putative dopaminergic neurons of the midbrain, and thus align
844 with our hypothesis that the neural signatures of working memory updating subprocess
845 substitution resemble updating cortical value representations through a (dopaminergic) reward
846 prediction error also via the basal ganglia (Montague et al., 1996; Schultz et al., 1997; Schultz,
847 2013), and thus the value-updating network associated with reinforcement learning. To what
848 extent these networks overlap precisely represents an exciting avenue for future research
849 employing methods such as joint-modeling of functional and behavioral data across tasks
850 (Palestro et al., 2018; Stevenson et al., 2024).

851 Finally, our findings indicate a role for dopaminergic brain regions in the substitution of
852 working memory rather than gating, even though it is unclear to what extent working memory
853 gating, as postulated in the PBWM, can be measured accurately by using the reference-back task
854 (see previous section 'No support for striatal gate opening'). Yet, our findings hint toward phasic
855 dopamine representing a signal to alter (in this framework, substitute) representations held in
856 working memory. This suggests that phasic dopamine release signals updating cortical
857 representations independent of the cognitive domain.

858 Taken together, our results provide convincing evidence for dopaminergic processing
859 when substituting new information into working memory. Moreover, our results suggest shared
860 mechanisms between working memory updating and value-based (reinforcement) learning, both
861 of which update cortical representations in an adaptive manner. Additionally, these findings
862 refine the PBWM model's gating mechanism and suggest that phasic dopamine release, in fact,
863 may signal the updating of cortical representations.

864 **Subcortical involvement in gate closing**

Based on the whole-brain and cluster-based ROI analyses, we found that the (left) PPC was the primary region involved in gate closing, without involvement from the basal ganglia. This finding is in line with Nir-Cohen et al. (2020) and the predictions of the PBWM model (Frank et al., 2001; Hazy et al., 2007; O'Reilly & Frank, 2006). However, it stands in contrast to Nir-Cohen et al. (2023), who found that the striatum was involved during distractor conditions that required active gate closing (by filtering out conflicting task cues). Intriguingly, our ROI-wise GLM analyses did reveal additional yet weak support for (right-lateralized) striatal activity during gate closing. Furthermore, our results provide weak evidence for lateralized right STN and SN activation and particularly strong evidence for the right GPe and right SNc activation during the process of gate closing. Across all fMRI analyses, we observed little activation of the PFC and large activation clusters in the PPC. However, given the evidence against thalamic activations, it seems unlikely that the entire basal ganglia-thalamo-cortical loop is engaged during gate closing. Thus, the ROI-wise GLM-based evidence implicating parts of the right basal ganglia in gate closing may hint toward a different functional basal ganglia loop. This is supported by evidence against the activity of the basal ganglia output nucleus GPi and the thalamus in this contrast, suggesting no engagement of the two structures crucial for the signaling between the cortex and basal ganglia.

It is possible that the neural signature of gate closing represents the suppression of inappropriate actions (i.e., working memory updating). Interestingly, the dominance of the right hemisphere in the subcortex observed in the ROI-wise GLMs contrasts with the predominantly left hemisphere findings from the whole-brain analysis. This difference in hemispheric dominance may suggest distinct lateralized functional roles in gate closing trials. On the one hand, inhibition tasks support a right-lateralized inhibitory network involving the subcortical regions striatum, GPe, and SN (Maizey et al., 2020; Isherwood et al., 2023). On the other hand, Maizey et al. (2020) observed minimal BOLD changes in the expected right cortical area, the inferior frontal gyrus, during active ignorance of updating, potentially explaining the relative prominence of left hemispheric cortical activations in our findings. This pattern is reminiscent of gate closing trials, possibly indicating an active suppression of updating working memory. In fact, a subset of GPe neurons –which exhibits the strongest evidence of activity in gate closing compared to any other subcortical ROI–projects to the striatum, where they powerfully suppress neurons of the direct and indirect pathway (Bevan, 2021; Glajch et al., 2016; Mallet et al., 2012, 2016). Together, this may indicate an inhibition process within the basal ganglia in our data–supported by the lack of behavioral costs during gate closing, in fact, increased accuracy observed akin to Boag et al. (2021) and Konjusha et al. (2023)–ensuring no erroneous updating signal is transmitted through the basal ganglia–thalamo–cortical loop.

Hence, our findings suggest that the basal ganglia circuit is not engaged in an active gate closing process due to the lack of evidence involving essential parts of the basal ganglia–thalamo–cortical loop, but instead supports preventing updating related signaling in the subcortex during trials requiring the protection of working memory content. As a result, findings imply the working memory gate is closed by default and that substituting working memory requires active signaling through the otherwise inhibited basal ganglia.

Finally, no evidence for activity in the VTA was found in gate closing, proving no support for gate closing via the mesocortical pathway, as suggested by Nir-Cohen et al. (2020) and Durstewitz and Seamans (2008). However, ROI-wise GLMs indicate a potential role of the nigrostriatal pathway originating in the SNc during gate closing, which is a surprising finding. It might represent an erroneous, premature updating signal that engages the subsequent GPe-driven inhibition; however, this remains entirely speculative.

Taken together, our analyses provide evidence against a striatal gate-closing process yet implicate the involvement of other subcortical nuclei (right GPe and the right SNc) during gate closing, which might represent a process to inhibit false updating-related signaling from being transmitted to the cortex. Furthermore, it supports the idea that the working memory gate is closed by default due to the basal ganglia sitting in a NoGo-state, and updating working memory requires actively transmitting a Go-signal through the basal ganglia.

## Functional subdivisions of the dopaminergic midbrain

Gate closing, substitution, and updating mode are differently associated with the activity of the individually parcellated midbrain ROIs VTA and SN. Yet, it remains unclear whether nigrostriatal, mesolimbic, or mesocortical cell populations were involved in the individual contrasts due to limitations in identifying the associated cell assemblies in the individual MRI scan. In the following, dopaminergic-midbrain-associated results across all ROI analyses are discussed in detail in order to potentially shed light on the functional spatial organization of the dopaminergic midbrain region.

For gate closing, in contrast to only very weak evidence from the individual masks by MASSP (Bazin et al., 2020), the use of the masks by Pauli et al. (2018) revealed bilateral involvement of the SNc in gate closing–particularly very strong evidence for the right SNc–implicating a role for the dopamine neurons of the SN in closing the gate to working memory. The right SNc was also engaged during substitution, in addition to the right PBP, as suggested by ROI-wise GLMs. Interestingly, in contrast to these ROI-wise GLM results for substitution, the cluster-based ROI analysis for substitution using the probabilistic VTA and SN masks from MASSP (Bazin et al., 2020) showed broad bilateral activation for both, while analyses using Pauli et al.'s (2018) masks implicated only the SNc, bilaterally (Tab. A2). Nonetheless, all three ROI analyses support the idea of dopaminergic modulation of working memory substitution. Overall, the cluster-based ROI analysis revealed activated clusters in the midbrain limited to the substitution contrast, with no activated midbrain clusters in updating mode. However, ROI-wise GLMs using the probabilistic atlas from Pauli et al. (2018) revealed strong evidence for the right SNc in updating mode, accompanied by evidence for the left SNc and right PBP. These findings suggest a pronounced involvement of a dopaminergic mechanism in this particular working memory updating mode. On the contrary, no evidence was observed for the VTA nucleus mask by Pauli et al. (2018) in any working memory subprocess, which may suggest that this mask does not cover neural populations in the VTA related to working memory updating.

These findings suggest initial evidence for nigrostriatal involvement in gate closing and dopaminergic influence in the control of substituting new information into working memory and a ready-to-update mode. However, the ROI-wise GLMs did not support the bilateral

engagement of the SNc in substitution indicated by the cluster-based ROI results and instead additionally implicated the right PBP nucleus. On the one hand, this highlights the differences in the two ROI analysis approaches, using unsmoothed and smoothed data, respectively, as discussed earlier. On the other hand, it challenges the regional specificity suggested if only cluster-based ROI results were considered: Given the directly adjacent location of the two nuclei, it is unclear whether exclusively nigrostriatal neurons are involved during working memory substitution. This ambiguity highlights the necessity of delineating PBP and SNc masks, preferably based on the A10 and A9 cell groups, respectively, in the individual. This is crucial for accurately distinguishing between contributions from the mesolimbic and nigrostriatal pathways (Haber, 2003; Haber & Knutson, 2010), especially on an individual basis.

Nevertheless, the strong evidence from various ROI analyses for SNc activity in gate closing, substitution, and updating mode clearly points towards the involvement of the nigrostriatal pathway and, thus, the basal ganglia–thalamo–cortical loop associated with motor function. Along these lines, there lies great potential in investigating the individual striatal subdivision to shed light on the contributions of individual dopaminergic pathways in working memory updating processes. The results from both whole-brain GLM and cluster-based ROI analyses indicate that the putamen and caudate are active during substitution and updating mode (please refer to the appendix, Tab. A1). However, it remains unclear how helpful individual parcellations-based ROI-wise GLMs would be in answering this question if they become available. After all, findings implicate the dorsal striatum, which comprises both the dorsal regions of the putamen and the caudate, which is classically associated with the nigrostriatal pathway (Haber, 2003; Haber & Knutson, 2010).

Furthermore, the findings associated with working memory substitution suggest meaningful differences between the VTA masks from the two ROI sources. Despite the equal amount of evidence for the VTA and PBP in substitution and updating mode, distinct functional differences associated with the underlying neural populations have become apparent. Specifically, the MASSP VTA mapping includes a large number of voxels that appear to be involved in working memory substitution, while Pauli et al.'s (2018) VTA nucleus mask did not elicit activation. Apparently, it does not functionally overlap with the MASSP VTA voxels involved in working memory substitution. The latter is further supported by the finding implicating the right PBP in substitution, reinforcing that the substitution-related activity is located in the dorsolateral VTA, where the PBP nucleus is situated (Trutti et al., 2019). This again underscores the demand for enhanced delineation of the VTA, particularly in distinguishing the dorsolateral VTA functionally from the ventromedial site.

To summarize, follow-up analyses using different masks associated with dopaminergic cell populations of the midbrain revealed interesting insights. There was strong evidence for the role of the SNc in the process of substitution, supported by both ROI analysis approaches that complement each other's limitations. Also, during gate closing and updating mode, the SNc showed increased activation, suggesting nigrostriatal pathway engagement. Additionally, there was evidence of the right PBP nucleus activity for both substitution and updating mode. At the same time, no contrast revealed evidence for working memory updating-related activity in the VTA nucleus. These observations suggest that the evidence for activity limited to the right VTA

using the MASSP masks (Bazin et al., 2020) was driven by activity in a region of the midbrain that is associated with both neighboring nuclei, PBP and SNc. The intricacy in the functional engagement by midbrain nuclei emphasizes the necessity for individually parcellated masks, in particular of the VTA, in future studies that provide more precision both with respect to functional and anatomical subdivisions. This becomes especially apparent when considering the likelihood of neighboring nuclei influencing the observed activity (de Hollander et al., 2017).

## Limitations

We replicated many behavioral results from previous reference-back studies (e.g., Boag et al., 2021; Jongkees, 2020; Rac-Lubashevky & Kessler, 2016a, 2016b, 2018). Notably, we observed an accuracy gain in gate closing trials similar to the findings of Boag et al. (2021) and Konjusha et al. (2023) but inconsistent with Nir-Cohen et al. (2020), indicating that an active switch of closing the gate potentially suppresses erroneous updating signal may be advantageous for working memory maintenance. However, we could not replicate the gate closing cost for response time that other studies have reported (e.g., Jongkees, 2020; Nir-Cohen et al., 2020; Rac-Lubashevky & Kessler, 2016a, 2016b, 2018). This may have been due to our ultra-high field scanning protocol, which allowed for fewer trials than are typically collected in reference-back studies and included longer inter-trial intervals than previous behavioral studies.

Consequently, the fewer trials compared to other work (Nir-Cohen et al., 2020, 2023) may have lowered the fMRI signal-to-noise ratio, potentially resulting in decreased efficacy in detecting neural markers associated with gate opening in the basal ganglia. Nevertheless, compared to other work, our protocol did unveil a substantially broader range of cortical and subcortical activation during other working memory processes, including the involvement of the right thalamus during gate opening, which is a critical part of the basal ganglia-thalamo-cortical loop. Consequently, it remains unclear if this accounts for the present findings. Future work employing ultra-high field fMRI using a higher trial-number design may shed light on more signals in the subcortical mechanisms involved in the process of gate opening.

Finally, the ROI-wise GLM analysis does not consider striatal subdivisions as these are not parcellated by MASSP. In light of the relatively large volume of the striatum as a whole, the association of the dorsal and ventral striatum with the nigrostriatal and mesolimbic pathways, respectively, in combination with the different functional profiles of the striatal subdivision (Haber, 2003), it would be intriguing to explore the contribution of individual striatal subdivision in working memory updating processes in the future, as discussed in the previous paragraph.

## Conclusions

Our finding of the absence of evidence for striatal activity during the process of gate opening, alongside observed activity in FPN regions, suggests the striatum is not crucially involved in opening the gate to working memory as operationalized in the reference-back task. This finding challenges the concept of a striatal working memory gating mechanism in such low-demand working memory tasks, and previous empirical results (Nir-Cohen et al., 2020, 2023). Instead, our data revealed basal ganglia engagement associated with the ready-to-update working memory mode ('updating mode'), suggesting a more specific role for the basal ganglia in a sustained ready-to-update mode rather than controlling the gate state to working memory.

Moreover, in line with previous work (Nir-Cohen et al., 2020), our findings implicate that the basal ganglia-thalamo-cortical loop is not engaged in gate closing, as evidence for basal ganglia output is lacking, despite observed basal ganglia activation. Instead, our ROI analyses, together with behavioral results, indicate the suppression of updating-related signaling in the subcortex during gate closing. Furthermore, our data indicate the involvement of dopamine-producing midbrain nuclei during the process of working memory gate closing, notably only the SNc, and during a general ready-to-update mode, both the SNc and the PBP of the VTA.

In addition, we found evidence suggesting that substituting new information into working memory is driven by dopaminergic midbrain nuclei–notably from all three; SNc, VTA, and PBP–via the basal ganglia and involving the basal ganglia-thalamo-cortical loop. This finding is consistent with neural signatures of value updating triggered by dopaminergic signals via reward prediction errors in reinforcement learning, potentially pointing to a common network underlying the updating of cortical representations which could be further quantified with joint-modeling in future studies (Stevenson et al., 2024). Additionally, connectivity and PET imaging could provide detailed insights to identify a specific dopaminergic pathway, enabling the quantification of the involvement of the mesolimbic and nigrostriatal pathways.

This study furthers understanding of the neural mechanisms underlying working memory updating subprocesses in the human subcortex, providing additional insights into the role of the dopaminergic midbrain. Moreover, it contributes to the development of methodologies for exploring the neurobiological basis of cognitive flexibility and adaptive behavior.

## References


Abraham, A., Pedregosa, F., Eickenberg, M., Gervais, P., Mueller, A., Kossaifi, J., Gramfort, A., Thirion, B. & Varoquaux, G. (2014). Machine learning for neuroimaging with scikit-learn. *Frontiers in Neuroinformatics*, *8*(2), 1–10. https://doi.org/10.3389/fninf.2014.00014

Armbruster, D. J., Ueltzhöffer, K., Basten, U., & Fiebach, C. J. (2012). Prefrontal cortical mechanisms underlying individual differences in cognitive flexibility and stability. *Journal of cCgnitive Neuroscience, 24*(12), 2385-2399. https://doi.org/10.1162/jocn_a_00286

Bazin, P.-L., Alkemade, A., Mulder, M. J., Henry, A. G.. & Forstmann, B. U. (2020). Multi-contrast anatomical subcortical structures parcellation. *eLife, 9*, 1–23. https://doi.org/10.7554/eLife.59430

Bäckman, L., Waris, O., Johansson, J., Andersson, M., Rinne, J. O., Alakurtti, K., Soveri, A., Laine, M., & Nyberg, L. (2017). Increased dopamine release after working-memory updating training: Neurochemical correlates of transfer. *Scientific Reports*, *7*(1). https://doi.org/10.1038/s41598-017-07577-y

Bedwell, J. S., Horner, M. D., Yamanaka, K., Li, X., Myrick, H., Nahas, Z., & George, M. S. (2005). Functional neuroanatomy of subcomponent cognitive processes involved in verbal working memory. *International Journal of Neuroscience, 115*(7), 1017-1032. https://doi.org/10.1080/00207450590901530

Behzadi, Y., Restom, K., Liau, J., & Liu, T. T. (2007). A component based noise correction method (CompCor) for BOLD and perfusion based fMRI. *NeuroImage, 37*(1), 90–101. https://doi.org/10.1016/j.neuroimage.2007.04.042

Bevan, M. D. (2021). Motor control: Basal ganglia feedback circuit for action suppression. *Current Biology, 31*(4), R191-R193. https://doi.org/10.1016/j.cub.2020.11.067

Birn, R. M., Smith, M. A., Jones, T. B., & Bandettini, P. A. (2008). The respiration response function: The temporal dynamics of fMRI signal fluctuations related to changes in respiration. *NeuroImage, 40*(2), 644–654. https://doi.org/10.1016/j.neuroimage.2007.11.059

Boag, R. J., Stevenson, N., van Dooren, R., Trutti, A. C., Sjoerds, Z., & Forstmann, B. U. (2021). Cognitive control of working memory: A model-based approach. *Brain Sciences*, *11*(6), 721. https://doi.org/10.3390/brainsci11060721

Braver, T. S., & Cohen, J. D. (2000). *On the control of control: The role of dopamine in regulating prefrontal function and working memory.* In S. Monsell & J. Drive (Eds.), Control of Cognitive Processes: Attention and Performance Xviii (pp. 712–737). https://doi.org/10.7551/mitpress/1481.001.0001

Chang, C., Crottaz-Herbette, S., & Menon, V. (2007). Temporal dynamics of basal ganglia response and connectivity during verbal working memory. *NeuroImage, 34*(3), 1253-1269. https://doi.org/10.1016/j.neuroimage.2006.08.056

Chang, C., Cunningham, J. P., & Glover, G. H. (2009). Influence of heart rate on the BOLD signal: The cardiac response function. *NeuroImage, 44*(3), 857–869. https://doi.org/10.1016/j.neuroimage.2008.09.029

Chatham, C. H., & Badre, D. (2015). Multiple gates on working memory. *Current Opinion in Behavioral Sciences, 1*, 23-31. https://doi.org/10.1016/j.cobeha.2014.08.001

Cohen, J. D., Braver, T. S., & Brown, J. W. (2002). Computational perspectives on dopamine function in prefrontal cortex. *Current Opinion In Neurobiology*, *12*(2), 223–229. https://doi.org/10.1016/s0959-4388(02)00314-8

Cools, R. (2006). Dopaminergic modulation of cognitive function-implications for l-DOPA treatment in Parkinson's disease. *Neuroscience & Biobehavioral Reviews, 30*(1), 1-23. https://doi.org/10.1016/j.neubiorev.2005.03.024

Cools, R., Clark, L., & Robbins, T. W. (2004). Differential responses in human striatum and prefrontal cortex to changes in object and rule relevance. *The Journal of Neuroscience, 24*(5), 1129–1135. https://doi.org/10.1523/JNEUROSCI.4312-03.2004

Cools, R. & D'Esposito, M. (2011). Inverted-U-shaped dopamine actions on human working memory and cognitive control. *Biological Psychiatry, 69*(12), https://doi.org/10.1016/j.biopsych.2011.03.028e113-e125.

Cooper, J. C., Dunne, S., Furey, T., & O'Doherty, J. P. (2012). Human dorsal striatum encodes prediction errors during observational learning of instrumental actions. *Journal of Cognitive Neuroscience, 24*(1), 106–118. https://doi.org/10.1162/jocn_a_00114

Corlett, P. R., Mollick, J. A., & Kober, H. (2022). Meta-analysis of human prediction error for incentives, perception, cognition, and action. *Neuropsychopharmacology, 47*(7), 1339–1349. https://doi.org/10.1038/s41386-021-01264-3

Cox, R., W., & Hyde, J. S. (1997). Software tools for analysis and visualization of fMRI data. *NMR in Biomedicine, 10* (4-5), 171–78. https://doi.org/10.1002/(sici)1099-1492(199706/08)10:4/5

D'Ardenne, K., Eshel, N., Luka, J., Lenartowicz, A., Nystrom, L. E., & Cohen, J. D. (2012). Role of prefrontal cortex and the midbrain dopamine system in working memory updating. *Proceedings of the National Academy of Sciences, 1109*(49), 19900-19909. https://doi.org/10.1073/pnas.1116727109

de Hollander, G., Keuken, M. C., & Forstmann, B. U. (2015). The Subcortical Cocktail Problem; Mixed Signals from the Subthalamic Nucleus and Substantia Nigra. *PloS One, 10*(3), e0120572. https://doi.org/10.1371/journal.pone.0120572

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/10.1016/j.cortex.2022.06.014

Dreisbach, G., (2012). Mechanisms of cognitive control: The Functional Role of Task Rules. *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

Durstewitz, D., & Seamans, J.K. (2008). The dual-state theory of prefrontal cortex dopamine function with relevance to catechol-O-methyltransferase genotypes and schizophrenia. *Biolological Psychiatry, 64*(9), 739-749. https://doi.org/10.1016/j.biopsych.2008.05.015

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), 170–189. https://doi.org/10.1037/a0017891

Esteban, O., Markiewicz, C. J., Blair, R. W., Moodie, C. A., Isik, A. I., Erramuzpe, A., Kent, J. D., Goncalves, M., DuPre, E., Snyder, M., Oya, H., Ghosh, S. S., Wright, J., Durnez, J., Poldrack, R. A., & Gorgolewski. K. J. (2019). fMRIPrep: a robust preprocessing pipeline for functional MRI. *Nature Methods, 16*(1), 111-116. https://doi.org/10.1038/s41592-018-0235-4

Esteban, O., Markiewicz, C. J., Goncalves, M., Provins, C., Salo, T., Kent, J. D., DuPre, E., Ciric, R., Pinsard, B., Blair, R. W., Poldrack, R. A., & Gorgolewski, K. J. (2018). fMRIPrep. Software. Zenodo. https://doi.org/10.5281/zenodo.852659

Fonov, V. S., Evans, A. C., McKinstry, R. C., Almli, C. R., & Collins, D. L. (2009). Unbiased nonlinear average age-appropriate brain templates from birth to adulthood. *NeuroImage, 47* (Supplement 1), S102. https://doi.org/10.1016/S1053-8119(09)70884-5.

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*(2), 137–160. https://doi.org/10.3758/cabn.1.2.137

Frässle, S., Aponte, E. A., Bollmann, S., Brodersen, K. H., Do, C. T., Harrison, O. K., Heinzle, J., Iglesias, S., Kasper, L., Lomakina, E. I., Mathys, C., Müller-Schrader, M., Pereira, I., Petzschner, F. H., Raman, S., Schöbi, D., Toussaint, B., Weber, L. A., Yao, Y., … & Stephan, K. E. (2021). TAPAS: An open-source software package for translational neuromodeling and computational psychiatry. *Frontiers in Psychiatry, 12,* 857. https://doi.org/10.3389/fpsyt.2021.680811

Glajch, K. E., Kelver, D. A., Hegeman, D. J., Cui, Q., Xenias, H. S., Augustine, E. C., Hernandez, V. M., Verma, N., Huang, T. Y., Luo, M., Justice, N. J. & Chan, C. S.et al. (2016). Npas1+ pallidal neurons target striatal projection neurons. *Journal of Neuroscience, 36*(20), 5472–5488. https://doi.org/10.1523/jneurosci.1720-15.2016

Glover, G. H. (1999). Deconvolution of impulse response in event-related BOLD fMRI1. *NeuroImage, 9*(4), 416–429. https://doi.org/10.1006/nimg.1998.0419

Glover, G. H., Li, T.-Q., & Ress, D. (2000). Image-based method for retrospective correction of physiological motion effects in fMRI: RETROICOR. *Magnetic Resonance in Medicine*, *44*(1), 162–167. https://doi.org/10.1002/1522-2594(200007)44:1%3C162::AID-MRM23%3E3.0.CO;2-E

Goldman-Rakic, P. (1995). Cellular basis of working memory. *Neuron, 14*(3)*,* 477-485. https://doi.org/10.1016/0896-6273(95)90304-6

Gorgolewski, K. J., Esteban, O., Markiewicz, C. J., Ziegler, E., Ellis, D. G., Notter, M. P., Jarecka, D., et al. (2018). Nipype. Software. Zenodo. https://doi.org/10.5281/zenodo.596855

Gorgolewski, K., Burns, C. D., Madison, C., Clark, D., Halchenko, Y. O., Waskom, M. L., & Ghosh, S. (2011). Nipype: A flexible, lightweight and extensible neuroimaging data processing framework in Python. *Frontiers in Neuroinformatics, 5*, 13. https://doi.org/10.3389/fninf.2011.00013

Goschke, T., & Bolte, A. (2014). Emotional modulation of control dilemmas: the role of positive affect, reward, and dopamine in cognitive stability and flexibility. *Neuropsychologia*, *62*, 403–423. https://doi.org/10.1016/j.neuropsychologia.2014.07.015

Greve, D. N., & Fischl, B. (2009). Accurate and robust brain image alignment using boundary-based registration. *NeuroImage, 48*(1), 63–72. https://doi.org/10.1016/j.neuroimage.2009.06.060.

Haber, S. N. (2003). The primate basal ganglia: parallel and integrative networks. *Journal of Chemical Neuroanatomy, 26*(4), 317-330. https://doi.org/10.1016/j.jchemneu.2003.10.003

Haber, S. N., & Knutson, B. (2010). The reward circuit: linking primate anatomy and human imaging. *Neuropsychopharmacology, 35*(1), 4-26. https://doi.org/10.1038/npp.2009.129

Harrison, S. J., Bianchi, S., Heinzle, J., Stephan, K. E., Iglesias, S., & Kasper, L. (2021). A Hilbert-based method for processing respiratory timeseries. *NeuroImage, 230*, 117787. https://doi.org/10.1016/j.neuroimage.2021.117787

Harvey, A. K., Pattinson, K. T., Brooks, J. C., Mayhew, S. D., Jenkinson, M., & Wise, R. G. (2008). Brainstem functional magnetic resonance imaging: Disentangling signal from physiological noise. *Journal of Magnetic Resonance Imaging, 28*(6), 1337–1344. https://doi.org/10.1002/jmri.21623

Hazy, T. E., Frank, M. J., & O'Reilly, R. C. (2007). Towards an executive without a homunculus: computational models of the prefrontal cortex/basal ganglia system. *Philosophical Transactions of he Royal Society of London. Series Biological sSiences*, *362*(1485), 1601–1613. https://doi.org/10.1098/rstb.2007.2055

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

Hommel, B. (2015). Chapter Two - Between persistence and flexibility: The yin and yang of action control. In A. J. Elliot (Ed.), *Advances in Motivation Science, Volume 2* (pp. 33-67).

Elsevier. ISSN 2215-0919, ISBN 9780128022702. https://doi.org/10.1016/bs.adms.2015.04.003

Isherwood, S. J. S., Bazin, Pl., Miletić, S., Stevenson, N. R., Trutti, A. C., Tse, D. H. Y., Heathcote, A., Matzke, D., Innes, R. J., Habli, S., Sokołowski, D. R., Alkemade, A., Håberg, A. K., & Forstmann, B. U. (2023). Investigating Intra-Individual Networks of Response Inhibition and Interference Resolution using 7T MRI. *NeuroImage*, *271*, 119988. https://doi.org/10.1016/j.neuroimage.2023.119988

Jacob, S. N., & Nieder, A. (2014). Complementary roles for primate frontal and parietal cortex in guarding working memory from distractor stimuli. *Neuron, 83*(1), 226-237. https://doi.org/10.1016/j.neuron.2014.05.009

Jeffreys, H., & Lindsay, R. B. (1939). *Theory of probability.* OuP Oxford.

Jenkinson, M., Bannister, P., Brady, M., & Smith, S. (2002). Improved optimization for the robust and accurate linear registration and motion correction of brain images. *NeuroImage, 17(*2), 825–41. https://doi.org/10.1006/nimg.2002.1132.

Jenkinson, M., Beckmann, C. F., Behrens, T. E., Woolrich, M. W., & Smith, S. M. (2012). FSL. *NeuroImage*, *62*(2), 782–790. https://doi.org/10.1016/j.neuroimage.2011.09.015

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. https://doi.org/10.1523/jneurosci.3904-10.2011

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.

Kasper, L., Bollmann, S., Diaconescu, A. O., Hutton, C., Heinzle, J., Iglesias, S., Hauser, T. U., Sebold, M., Manjaly, Z. M., Pruessmann, K. P., & Stephan, K. E. (2017). The PhysIO Toolbox for modeling physiological noise in fMRI data. *Journal of Neuroscience Methods, 276,* 56–72. https://doi.org/10.1016/j.jneumeth.2016.10.019

Konjusha, A., Yu, S., Mückschel, M., Colzato, L., Ziemssen, T., & Beste, C. (2023). Auricular transcutaneous vagus nerve stimulation specifically enhances working memory gate closing mechanism: a system neurophysiological study. *Journal Of Neuroscience, 43*(25), 4709–4724. https://doi.org/10.1523/jneurosci.2004-22.2023

Lanczos, C. (1964). Evaluation of noisy data. *Journal of the Society for Industrial and Applied Mathematics Series B Numerical Analysis, 1*(1), 76–85. https://doi.org/10.1137/0701007

Leber, A. B., Turk-Browne, N. B., & Chun, M. M. (2008). Neural predictors of within-subject fluctuations in attentional control. *Journal of Neuroscience, 30*(34), 11458-11465. https://doi.org/10.1523/jneurosci.0809-10.2010

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. https://doi.org/10.1111/nyas.13714

Maizey, L., Evans, C. J., Muhlert, N., Verbruggen, F., Chambers, C. D., & Allen, C. P. (2020). Cortical and subcortical functional specificity associated with response inhibition. *NeuroImage, 220*, 117110. https://doi.org/10.1016/j.neuroimage.2020.117110

Mallet, N., Micklem, B. R., Henny, P., Brown, M. T., Williams, C., Bolam, J. P., Nakamura, K. C., & Magill, P. J. (2012). Dichotomous organization of the external globus pallidus. *Neuron, 74*(6), 1075–1086. https://doi.org/10.1016/j.neuron.2012.04.027

Mallet, N., Schmidt, R., Leventhal, D., Chen, F., Amer, N., Boraud, T., & Berke, J.D. (2016). Arkypallidal cells send a stop signal to striatum. *Neuron, 89*(2), 308–316. https://doi.org/10.1016/j.neuron.2015.12.017

Marques, J. P., Kober, T., Krueger, G., van der Zwaag, W., van de Moortele, P. F., & Gruetter, R. (2010). MP2RAGE, a self bias-field corrected sequence for improved segmentation and T1-mapping at high field. *NeuroImage, 49*(2), 1271–1281. https://doi.org/10.1016/j.neuroimage.2009.10.002

Marvel, C. L., Morgan, O. P., & Kronemer, S. I. (2019). How the motor system integrates with working memory. *Neuroscience & Biobehavioral Reviews, 102*, 184-194. https://doi.org/10.1016/j.neubiorev.2019.04.017

Mehta, M. A., Owen, A. M., Sahakian, B. J., Mavaddat, N., Pickard, J. D., & Robbins, T. W. (2000). Methylphenidate enhances working memory by modulating discrete frontal and parietal lobe regions in the human brain. *The Journal of Neuroscience, 20*(6), 1-6. https://doi.org/10.1523/jneurosci.20-06-j0004.2000

Miletić, S. (2023). *Modelling structure and function of the human subcortex*. [Thesis, fully internal, Universiteit van Amsterdam].

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 7T. *NeuroImage, 219*, 116992. https://doi.org/10.1016/j.neuroimage.2020.116992

Miletić, S., Boag, R. J., Trutti, A. C., Stevenson, N., Forstmann, B. U., & Heathcote, A. (2021). A new model of decision processing in instrumental learning tasks. *eLife, 10*. https://doi.org/10.7554/eLife.63055

Montague, P. R., Dayan, P., & Sejnowski, T. J. (1996). A framework for mesencephalic dopamine systems based on predictive Hebbian learning. *Journal of Neuroscience, 16(*5), 1936-1947. https://doi.org/10.1523/JNEUROSCI.16-05-01936.1996

Moustafa, A. A., Sherman, S. J., & Frank, M. J. (2008). A dopaminergic basis for working memory, learning and attentional shifting in Parkinsonism. *Neuropsychologia, 46*(13), 3144-3156. https://doi.org/10.1016/j.neuropsychologia.2008.07.011

Murty, V. P., Sambataro, F., Radulescu, E., Altamura, M., Iudicello, J., Zoltick, B., Weinberger, D. R., Goldberg, T. E., & Mattay, V. S. (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

Nichols, T., & Hayasaka, S. (2003). Controlling the familywise error rate in functional neuroimaging: a comparative review. *Statistical Methods in Medical Research*, *12*(5), 419–446. https://doi.org/10.1191/0962280203sm341ra

Nir-Cohen, G., Egner, T., & Kessler, Y. (2023). The neural correlates of updating and gating in procedural working memory. *Journal of Cognitive Neuroscience, 35*(6), 919–940. https://doi.org/10.1162/jocn_a_01988

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

O'Doherty, J. P., Cockburn, J., & Pauli, W. M. (2017). Learning, reward, and decision making. *Annual Review of Psychology, 68*(1), 73-100. https://doi.org/10.1146/annurev-psych-010416-044216

O'Reilly, R. C. (2006). Biologically based computational models of high-level cognition. *Science, 314*(5796), 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

Oberauer, K. (2009). Design for a working memory. *Psychology of learning and motivation*, *51*, 45-100.

Ott, T., & Nieder, A. (2019). Dopamine and cognitive control in prefrontal cortex. *Trends in Cognitive Sciences, 23*(3), 213–234. https://doi.org/10.1016/j.tics.2018.12.006

Palestro, J. J., Bahg, G., Sederberg, P. B., Lu, Z. L., Steyvers, M., & Turner, B. M. (2018). A tutorial on joint models of neural and behavioral measures of cognition. *Journal of Mathematical Psychology*, *84*, 20-48. https://doi.org/10.1016/j.jmp.2018.03.003

Pauli, W., Nili, A., & Tyszka, J. (2018). A high-resolution probabilistic in vivo atlas of human subcortical brain nuclei. *Scientific Data, 5*(1), 1-13. https://doi.org/10.1038/sdata.2018.63

Pijnenburg, R., Scholtens, L. H., Ardesch, D. J., de Lange, S. C., Wei, Y., & van den Heuvel, M. P. (2021). Myelo- and cytoarchitectonic microstructural and functional human cortical atlases reconstructed in common MRI space. *NeuroImage, 239*, 118274. https://doi.org/10.1016/j.neuroimage.2021.118274

Ponzi, A. (2008, September). Dynamical model of action reinforcement by gated working memory. In Advances in Cognitive Neurodynamics ICCN 2007: Proceedings of the International Conference on Cognitive Neurodynamics. ICCN 2007 Proceedings (pp. 463-467). Dordrecht: Springer Netherlands.

Power, J. D., Mitra, A., Laumann, T. O., Snyder, A. Z., Schlaggar, B. L., & Petersen, S. E. (2014). Methods to detect, characterize, and remove motion artifact in resting state fMRI. *NeuroImage, 84*, (Supplement C), 320–41. https://doi.org/10.1016/j.neuroimage.2013.08.048

Rac-Lubashevsky, R., & Kessler, Y. (2016a). 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. (2016b). 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. (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

Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian t tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review, 16*, 225-237.

Satterthwaite, T. D., Elliott, M. A., Gerraty, R. T., Ruparel, K., Loughead, J., Calkins, M. E., Eickhoff, S. B., Hakonarson, H., Gur, R.C., Gur, R. E., & Wolf, D. H. (2013). An improved framework for confound regression and filtering for control of motion artifact in the preprocessing of resting-state functional connectivity data. *NeuroImage, 64*(1), 240–256. https://doi.org/10.1016/j.neuroimage.2012.08.052

Schridde, U., Khubchandani, M., Motelow, J. E., Sanganahalli, B. G., Hyder, F., & Blumenfeld, H. (2008). Negative BOLD with large increases in neuronal activity. *Cerebral cortex, 18*(8), 1814–1827. https://doi.org/10.1093/cercor/bhm208

Schultz, W. (2013). Updating dopamine reward signals. *Current Opinion in Neurobiology, 23*(2), 229-238. https://doi.org/10.1016/j.conb.2012.11.012

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

Smith, S., & J. Brady (1997). SUSAN—A new approach to low level image processing. *International Journal of Computer Vision, 23*,(1), 45–78. https://doi.org/10.1023/A:1007963824710

Stevenson, N., Innes, R., Boag, R. J., Miletic, S., Isherwood, S. J. S., Trutti, A. C., Heathcote, A., & Forstmann, B. U. (2024). Joint modelling of latent cognitive mechanisms shared across decision-making domains. *Computational Brain & Behavior, 7*, 1-22. https://doi.org/10.1007/s42113-023-00192-3

Sutton, R. S., & Barto, A. G. (2018). Reinforcement learning: An introduction. MIT press.

Trutti, A. C., Fontanesi, L., Mulder, M. J., Bazin, P.-L., Hommel, B., & Forstmann, B. U. (2021). A probabilistic atlas of the human ventral tegmental area (VTA) based on 7 Tesla MRI data. *Brain Structure and Function, 226*(4), 1155–1167. https://doi.org/10.1007/s00429-021-02231-w

Trutti, A. C., Mulder, M. J., Hommel, B., & Forstmann, B. U. (2019). Functional neuroanatomical review of the ventral tegmental area. *NeuroImage, 191*, 258-268. https://doi.org/10.1016/j.neuroimage.2019.01.062

Tustison, N. J., Avants, B. B., Cook, P. A., Zheng, Y., Egan, A., Yushkevich, P. A., & Gee, J. C. (2010). N4ITK: Improved N3 bias correction. *IEEE Transactions on Medical Imaging, 29*(6), 1310–20. https://doi.org/10.1109/TMI.2010.2046908

Vallentin, D., Bongard, S., & Nieder, A. (2012). Numerical rule coding in the prefrontal, premotor, and posterior parietal cortices of macaques. *Journal of Neuroscience, 32*(19), 6621-6630. https://doi.org/10.1523/JNEUROSCI.5071-11.2012

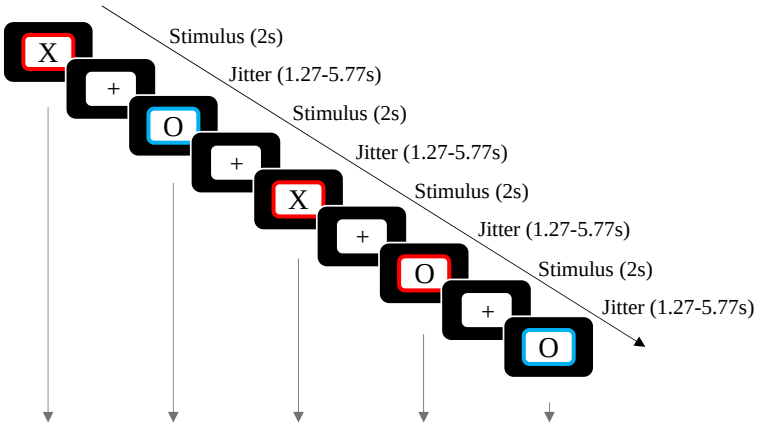
van Schouwenburg, M. R., Onnink, A. M. H., ter Huurne, N., Kan, C. C., Zwiers, M. P., Hoogman, M., Franke, B., Buitelaar, J. K., & Cools, R. (2014). Cognitive flexibility depends on white matter microstructure of the basal ganglia. *Neuropsychologia, 53*, 171-177. https://doi.org/10.1016/j.neuropsychologia.2013.11.015

Wade, A. R. (2002). The negative BOLD signal unmasked. *Neuron, 36*(6), 993–995. https://doi.org/10.1016/S0896-6273(02)01138-8

Woolrich, M. W., Ripley, B. D., Brady, M., & Smith, S. M. (2001). Temporal autocorrelation in univariate linear modeling of FMRI data. *NeuroImage, 14*(6), 1370–1386. https://doi.org/10.1006/nimg.2001.0931

Yekutieli, D., & Benjamini, Y. (1999). Resampling-based false discovery rate controlling multiple test procedures for correlated test statistics. *Journal of Statistical Planning and Inference, 82*(1-2), 171–196. https://doi.org/10.1016/S0378-3758(99)00041-5

Zhang, Y., Brady, M., & Smith, S. (2001). Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm. *IEEE Transactions on Medical Imaging, 20*(1), 45–57. https://doi.org/10.1109/42.906424




| | | | | | |
|---|---|---|---|---|---|
| **Trial Type** | reference | comparison | reference | reference | comparison |
| **Gate** | open | closed | open | open | closed |
| **Gate Switching** | | switch (closing) | switch (opening) | repeat | switch (closing) |
| **Response** | | different | same | different | same |

**Figure 1.** Example of the reference-back paradigm, including fMRI timing, as used in the experiment. In each trial, participants are tasked with indicating whether the probe stimulus ('X' or 'O') matches or differs from the stimulus presented in the most recent red frame, which serves as the working memory referent. In reference trials (red frame), participants are required to update their working memory with the currently displayed item. On the other hand, in comparison trials (blue frame), participants make the 'same/different' decision but do not update their working memory.

| **Trial type** | reference | | | | comparison | | | |
|---|---|---|---|---|---|---|---|---|
| **Switch type** | repeat | | switch | | repeat | | switch | |
| **Response type** | same | different | same | different | same | different | same | different |
| **Contrast** | | | | | | | | |
| *Gate opening* | - | - | + | + | | | | |
| *Gate closing* | | | | | - | - | + | + |
| *Substitution* | - | + | | | + | - | | |
| *Updating mode* | + | + | | | - | - | | |

**Figure 2.** Contrast weights for defining the four distinct working memory updating subprocesses.

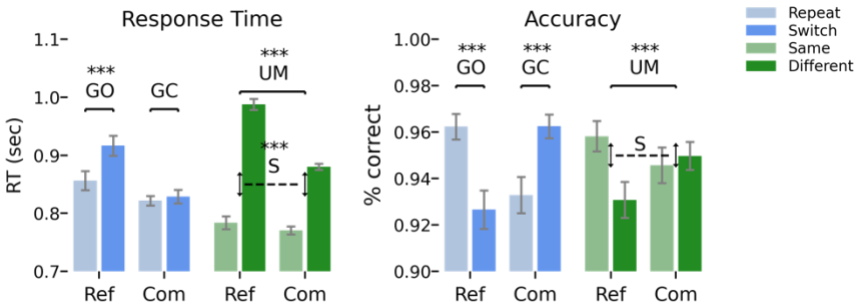



**Figure 3.** The figure illustrates the mean response times (and standard error of the mean) in relation to two factors: a) whether the condition was switched or repeated, and b) whether the stimulus/response matched the previous reference stimulus or differed from it. The figure also displays the associated behavioral contrasts and their significance levels. See 2.2. for detailed information on how the contrasts were computed. Ref = reference trials; Com = comparison trials; GO = gate opening; GC = gate closing; S = substitution; UM = updating mode. *** = $p < 0.001$

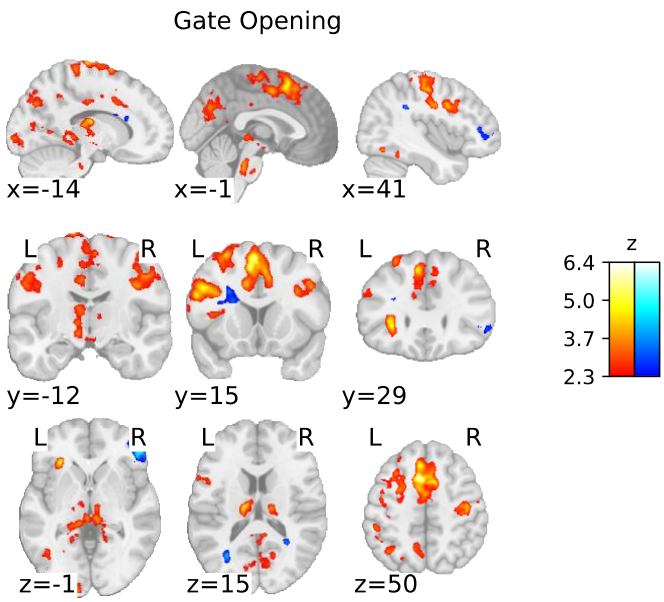



**Figure 4.** Statistical parametric map of the 'gate opening' contrast with a threshold determined using the FWER method ($p$<.05; corresponding to a threshold of z = 2.3).

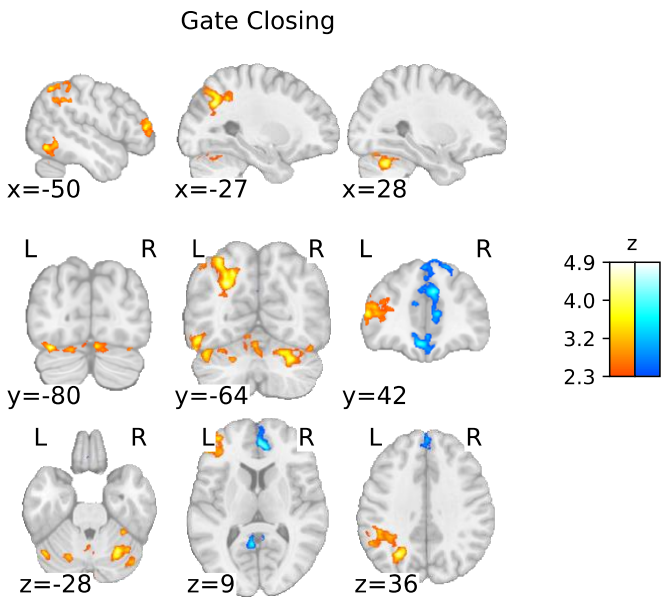



**Figure 5.** Statistical parametric map of the 'gate closing' contrast with a threshold determined using the FWER method ($p$<.05; corresponding to a threshold of z = 2.3).

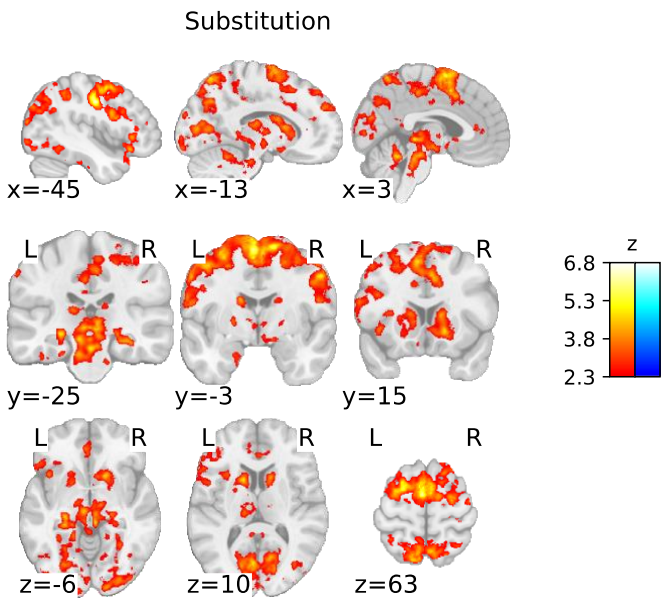



**Figure 6.** Statistical parametric map of the 'substitution' contrast with a threshold determined using the FWER method ($p$<.05; corresponding to a threshold of z = 2.3).

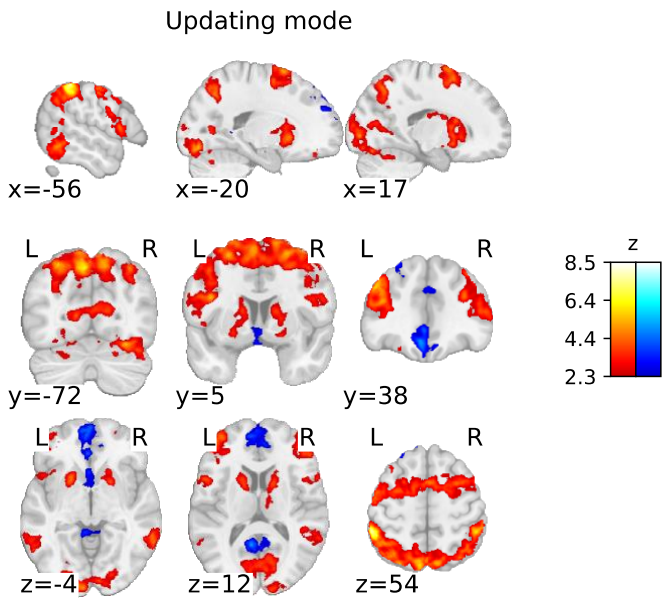



**Figure 7.** Statistical parametric map of the 'updating mode' contrast with a threshold determined using the FWER method ($p$<.05; corresponding to a threshold of z = 2.3).

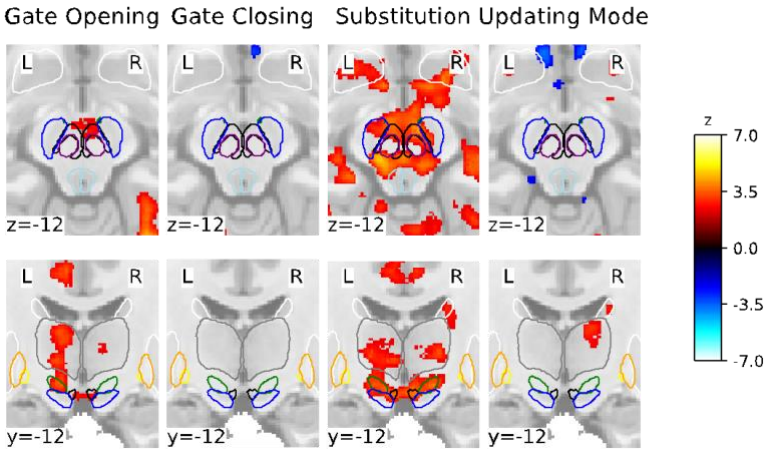



**Figure 8.** Subcortical statistical parametric mapping of the four contrasts with a focus on the midbrain (top row), and midbrain and thalamic regions (bottom row) using a FWER threshold of 2.3. Coordinates are in MNI2009cAsym (1mm) space. For orientation purposes, the MASSP atlas is overlaid: SN (blue), VTA (black), red nucleus (purple), subthalamic nucleus (green), periaqueductal grey (lightblue), GPi (yellow), GPe (orange), striatum (white), and thalamus (grey).

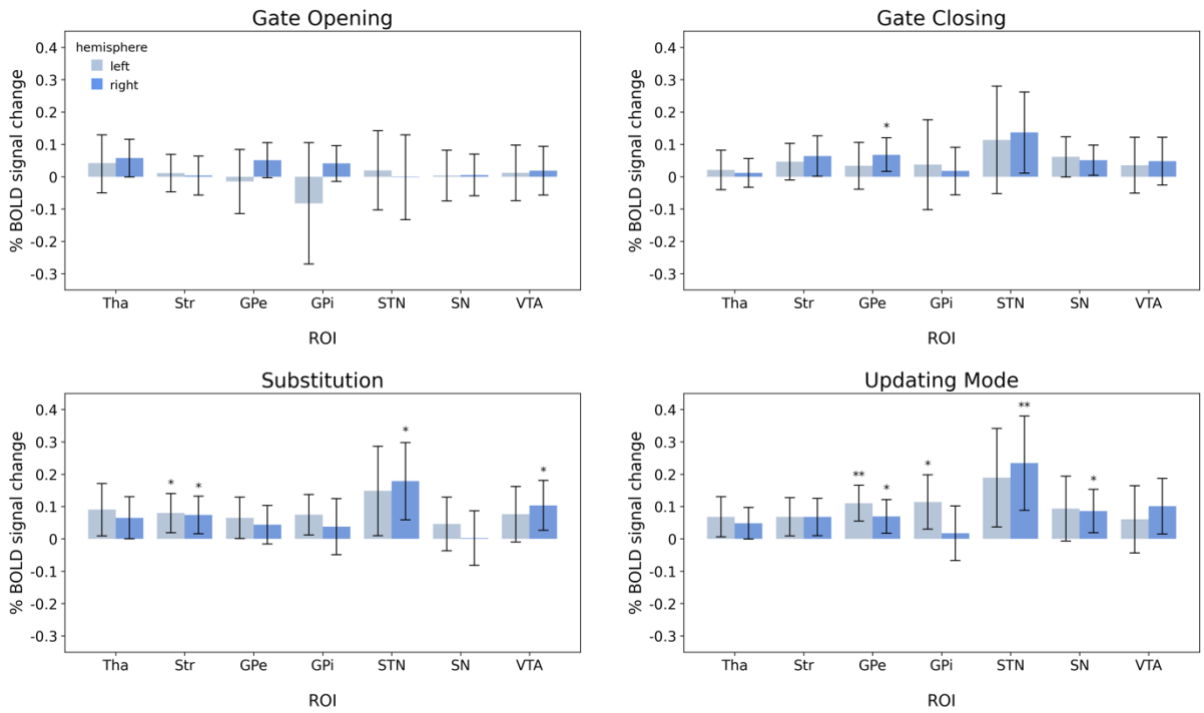



**Figure 9.** Results of the region-of-interest wise GLMs using the individually parcellated masks derived from MASSP. Error bars represent the 95% credible intervals. Abbreviations indicate thalamus (Tha), striatum (Str), globus pallidus externa (GPe), globus pallidus interna (GPi), subthalamic nucleus (STN), substantia nigra (SN), and ventral tegmental area (VTA). * indicate moderate evidence and ** indicate strong evidence that the observed BOLD signal change is larger than 0.

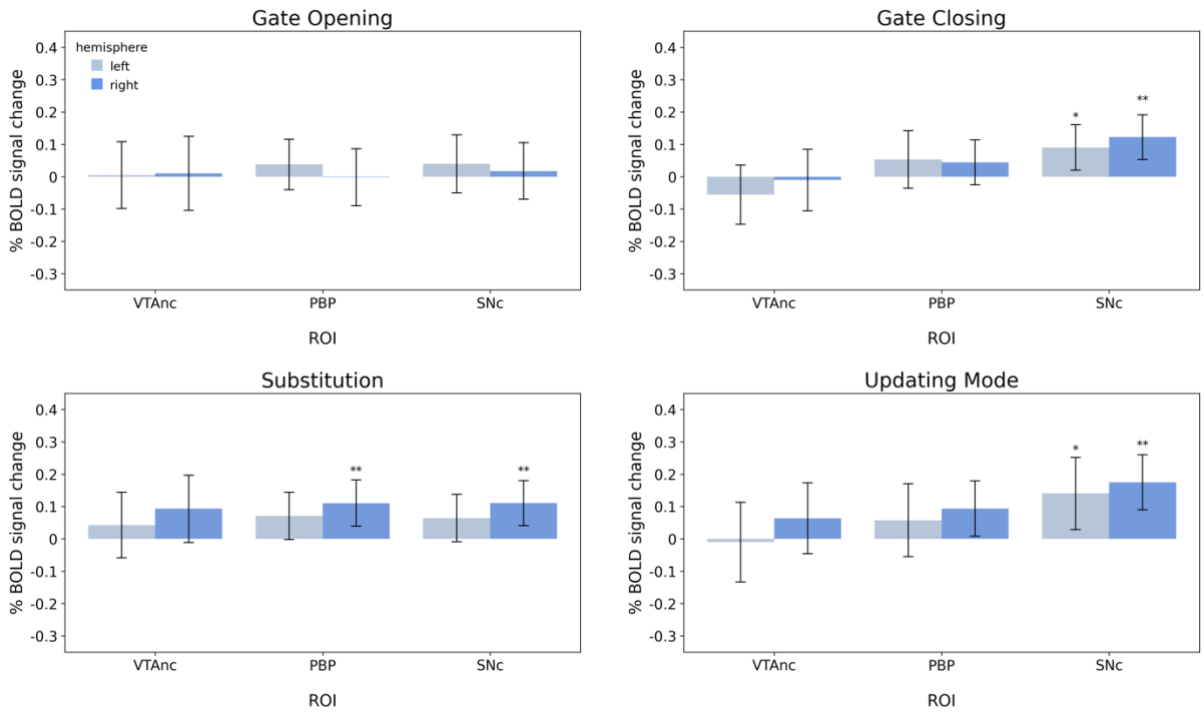



**Figure 10.** Results of the region-of-interest wise GLMs using the probabilistic atlas from Pauli et al. (2018). Error bars represent the 95% credible intervals. Abbreviations indicate ventral tegmental area nucleus (VTA), parabrachial pigmented nucleus (PBP), and substantia nigra pars compacta (SNc). * indicate moderate evidence and ** indicate strong evidence that the observed BOLD signal change is larger than 0.

# Tables

**Table 1.** Hierarchical descriptive statistics for the behavioral data.

| | | | Response Time | | Accuracy | |
|---|---|---|---|---|---|---|
| **Trial type** | **Switch type** | **Response type** | ***Mean (sec)*** | ***SD*** | ***Mean (%)*** | ***SD*** |
| Reference | Repeat | Same | 0.751 | 0.243 | 97.205 | 0.164 |
| | | Different | 0.957 | 0.269 | 95.298 | 0.211 |
| | Switch | Same | 0.807 | 0.274 | 94.560 | 0.227 |
| | | Different | 1.012 | 0.291 | 91.001 | 0.286 |
| Comparison | Repeat | Same | 0.776 | 0.262 | 93.085 | 0.254 |
| | | Different | 0.861 | 0.253 | 93.693 | 0.243 |
| | Switch | Same | 0.76 | 0.216 | 96.196 | 0.191 |
| | | Different | 0.896 | 0.255 | 96.357 | 0.187 |

**Table 2.** List of peak activation in MNI coordinates from the whole-brain analysis.

| | | | | MNI | | | |
|---|---|---|---|---|---|---|---|
| | | **hem** | **voxels** | **x** | **y** | **z** | **Z** |
| *Gate opening* | | | | | | | |
| | preSMA | l | 13380 | -6.0 | 12.0 | 50.0 | 6.387 |
| | Occipital fusiform gyrus | l | 8915 | -29.0 | -63.0 | -7.0 | 5.144 |
| | MFG | r | 849 | 34.0 | 9.0 | 29.0 | 5.129 |
| | Insular | l | 724 | -30.0 | 29.0 | -1.0 | 5.025 |
| | Precuneus cortex | l | 5482 | 7.0 | -67.0 | 25.0 | 4.422 |
| | Brainstem | | 841 | 0.0 | -33.0 | -29.0 | 4.338 |
| | Primary somatosensory cortex | r | 2043 | 41.0 | -16.0 | 50.0 | 4.219 |
| | M1 | l | 976 | -27.0 | -28.0 | 64.0 | 4.064 |
| | Primary somatosensory cortex | l | 566 | -53.0 | -14.0 | 40.0 | 3.958 |
| | M1 | r | 81 | 11.0 | -24.0 | 81.0 | 3.363 |
| | Primary somatosensory cortex | r | 34 | 66.0 | -8.0 | 30.0 | 3.284 |
| | Primary motor cortex | l | 92 | -35.0 | -20.0 | 45.0 | 3.188 |
| *Gate closing* | | | | | | | |
| | Inferior parietal cortex | l | 3769 | -27.0 | -64.0 | 36.0 | 4.560 |
| | Inferior temporal gyrus | l | 716 | -52.0 | -65.0 | -12.0 | 4.376 |
| | VI (cerebellum) | r | 2418 | 28.0 | -61.0 | -28.0 | 4.316 |
| | Fusiform gyrus | l | 957 | -33.0 | -80.0 | -20.0 | 4.284 |
| | V (cerebellum) | l | 1680 | -50.0 | 42.0 | 9.0 | 4.091 |
| | Fusiform gyrus | r | 14 | 35.0 | -79.0 | -21.0 | 3.070 |
| *Substitution* | | | | | | | |
| | M1 | l | 68499 | -45.0 | -13.0 | 38.0 | 6.834 |
| | Putamen | l | 1813 | -20.0 | 5.0 | 0.0 | 5.530 |
| | BA9 (dlPFC) | l | 714 | -22.0 | 56.0 | 33.0 | 4.763 |
| | Insular | l | 108 | -31.0 | 17.0 | 10.0 | 3.389 |
| | Brainstem | | 32 | -19.0 | -32.0 | -33.0 | 3.184 |
| | M1 | r | 15 | 25.0 | -25.0 | 68.0 | 2.974 |
| | BA9 (dlPFC) | l | 21 | -31.0 | 41.0 | 45.0 | 2.892 |
| | Parahippocampal Gyrus | l | 12 | -28.0 | -30.0 | -20.0 | 2.825 |
| | Visual cortex V1 | r | 12 | 29.0 | -62.0 | 5.0 | 2.803 |
| | Cingulate gyrus | r | 15 | 5.0 | -51.0 | 19.0 | 2.728 |
| | Inferior parietal lobule | l | 13 | -57.0 | -44.0 | 34.0 | 2.652 |
| | Visual cortex V4 | r | 19 | 22.0 | -71.0 | -1.0 | 2.633 |
| *Updating mode* | BA40 (PPC) | l | 24635 | -56.0 | -41.0 | 55.0 | 8.084 |
| | MFG | l | 19609 | -36.0 | 0.0 | 65.0 | 6.160 |
| | BA9 (dlPFC) | l | 5254 | -44.0 | 38.0 | 33.0 | 5.816 |
| | Fusiform gyrus | r | 12503 | 30.0 | -70.0 | -17.0 | 5.776 |
| | Putamen | l | 1486 | -21.0 | 15.0 | -2.0 | 4.983 |
| | Caudate | r | 1702 | 17.0 | 13.0 | 16.0 | 4.384 |
| | Inferior temporal gyrus | r | 45 | 49.0 | -38.0 | -16.0 | 3.162 |
| | Precuneus cortex | r | 19 | 18.0 | -60.0 | 31.0 | 2.754 |

**Table 3.** Results from the Bayesian one-sample *t*-test on the beta values derived from the ROI-wise GLMs are reported for each contrast and ROI. Moderate or higher evidence for ROI activity is indicated by bold font. Bayes factors favoring the alternative hypothesis are reported under $BF_{10}$, and Bayes factors favoring the null hypothesis in column $BF_{01}$.

| | ROI | hem | $BF_{10}$ | $BF_{01}$ | error % |
|---|---|---|---|---|---|
| *Gate Opening* | Tha | l | 0.287 | 3.489 | 0.037 |
| | | r | 1.070 | 0.934 | 0.023 |
| | Str | l | 0.197 | 5.072 | 0.041 |
| | | r | 0.185 | 5.407 | 0.041 |
| | GPe | l | 0.192 | 5.213 | 0.041 |
| | | r | 0.952 | 1.050 | 0.024 |
| | GPi | l | 0.266 | 3.755 | 0.038 |
| | | r | 0.510 | 1.962 | 0.031 |
| | STN | l | 0.193 | 5.184 | 0.041 |
| | | r | 0.184 | 5.440 | 0.041 |
| | SN | l | 0.184 | 5.421 | 0.041 |
| | | r | 0.186 | 5.372 | 0.041 |
| | VTA | l | 0.190 | 5.250 | 0.041 |
| | | r | 0.207 | 4.831 | 0.040 |
| *Gate Closing* | Tha | l | 0.229 | 4.364 | 0.039 |
| | | r | 0.208 | 4.799 | 0.040 |
| | Str | l | 0.643 | 1.556 | 0.028 |
| | | r | 1.239 | 0.807 | 0.021 |
| | GPe | l | 0.275 | 3.636 | 0.038 |
| | | r | **3.582** | **0.279** | **$6.359\times10^{-7}$** |
| | GPi | l | 0.211 | 4.746 | 0.040 |
| | | r | 0.205 | 4.874 | 0.040 |
| | STN | l | 0.444 | 2.255 | 0.033 |
| | | r | 1.568 | 0.638 | 0.019 |
| | SN | l | 1.104 | 0.905 | 0.022 |
| | | r | 1.537 | 0.651 | 0.019 |
| | VTA | l | 0.254 | 3.936 | 0.038 |
| | | r | 0.404 | 2.478 | 0.034 |
| *Substitution* | Tha | l | 1.722 | 0.581 | 0.018 |
| | | r | 1.138 | 0.879 | 0.022 |
| | Str | l | **3.846** | **0.260** | **$5.752\times10^{-7}$** |
| | | r | **3.292** | **0.304** | **$7.157\times10^{-7}$** |
| | GPe | l | 1.232 | 0.812 | 0.021 |
| | | r | 0.502 | 1.992 | 0.031 |
| | GPi | l | 2.367 | 0.422 | $1.108\times10^{-6}$ |
| | | r | 0.264 | 3.791 | 0.038 |
| | STN | l | 1.486 | 0.673 | 0.019 |
| | | r | **8.206** | **0.122** | **$1.685\times10^{-7}$** |
| | SN | l | 0.326 | 3.065 | 0.036 |
| | | r | 0.184 | 5.434 | 0.041 |
| | VTA | l | 0.779 | 1.284 | 0.026 |
| | | r | **4.089** | **0.245** | **$5.267\times10^{-7}$** |
| *Updating mode* | Tha | l | 1.605 | 0.623 | 0.019 |
| | | r | 1.035 | 0.967 | 0.023 |
| | Str | l | 1.943 | 0.515 | $1.415\times10^{-6}$ |
| | | r | 2.233 | 0.448 | $1.192\times10^{-6}$ |
| | GPe | l | **90.327** | **0.011** | **$1.536\times10^{-8}$** |
| | | r | **3.956** | **0.253** | **$5.523\times10^{-7}$** |
| | GPi | l | **4.451** | **0.225** | **$4.633\times10^{-7}$** |
| | | r | 0.199 | 5.024 | 0.040 |
| | STN | l | 2.862 | 0.349 | $8.649\times10^{-7}$ |
| | | r | **14.117** | **0.071** | **$5.350\times10^{-8}$** |
| | SN | l | 0.908 | 1.101 | 0.025 |
| | | r | **3.273** | **0.306** | **$7.215\times10^{-7}$** |
| | VTA | l | 0.348 | 2.877 | 0.035 |
| | | r | 2.185 | 0.458 | $1.225\times10^{-6}$ |

*Note.* For all tests, the alternative hypothesis specifies that the population mean differs from 0. $BF_{10}$ 1-3 indicates weak evidence, $BF_{10}$ 3-10 moderate evidence and $BF_{10}$ >10 strong evidence.

**Table 4.** Results from the Bayesian one-sample *t*-test on the beta values derived from the ROI GLMs using the masks from Pauli et al. (2018) are reported for each contrast and ROI. Moderate or higher evidence for ROI activity is indicated by bold font. Bayes factors favoring the alternative hypothesis are reported under $BF_{10}$, and Bayes factors favoring the null hypothesis in column $BF_{01}$.

| | ROI | hem | $BF_{10}$ | $BF_{01}$ | error % |
|---|---|---|---|---|---|
| *Gate opening* | VTAnc | l | 0.190 | 5.256 | 0.041 |
| | | r | 0.184 | 5.436 | 0.041 |
| | PBP | l | 0.351 | 2.851 | 0.035 |
| | | r | 0.185 | 5.392 | 0.041 |
| | SNc | l | 0.267 | 3.741 | 0.038 |
| | | r | 0.198 | 5.048 | 0.041 |
| *Gate closing* | VTAnc | l | 0.388 | 2.578 | 0.034 |
| | | r | 0.213 | 4.705 | 0.040 |
| | PBP | l | 0.340 | 2.944 | 0.035 |
| | | r | 0.414 | 2.415 | 0.033 |
| | SNc | l | **3.348** | **0.299** | **$6.990\times10^{-7}$** |
| | | r | **32.293** | **0.031** | **$4.835\times10^{-9}$** |
| *Substitution* | VTAnc | l | 0.184 | 5.429 | 0.041 |
| | | r | 0.459 | 2.178 | 0.032 |
| | PBP | l | 0.774 | 1.292 | 0.026 |
| | | r | **3.913** | **0.256** | **$5.609\times10^{-7}$** |
| | SNc | l | 0.754 | 1.326 | 0.027 |
| | | r | **13.330** | **0.075** | **$6.138\times10^{-8}$** |
| *Updating mode* | VTAnc | l | 0.224 | 4.474 | 0.040 |
| | | r | 0.343 | 2.916 | 0.035 |
| | PBP | l | 0.265 | 3.777 | 0.038 |
| | | r | 1.304 | 0.767 | 0.021 |
| | SNc | l | **3.105** | **0.322** | **$7.757\times10^{-7}$** |
| | | r | **142.529** | **0.007** | **$2.305\times10^{-4}$** |

*Note.* For all tests, the alternative hypothesis specifies that the population mean differs from 0. $BF_{10}$ 1-3 indicates weak evidence, $BF_{10}$ 3-10 moderate evidence and $BF_{10}$ >10 strong evidence.