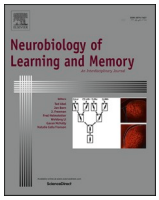
Neurobiology of Learning and Memory 179 (2021) 107387

Contents lists available at ScienceDirect

Neurobiology of Learning and Memory

journal homepage: www.elsevier.com/locate/ynlme



Consequence of stroke for feature recall and binding in visual working memory

Selma Lugtmeijer [a,b,*,1], Sebastian Schneegans [c,1], Nikki A. Lammers [a], Linda Geerligs [b], Frank Erik de Leeuw [d], Edward H.F. de Haan [a], Paul M. Bays [c], Roy P.C. Kessels [b,e]

[a] University of Amsterdam, Amsterdam, the Netherlands
[b] Radboud University, Donders Institute for Brain, Cognition and Behaviour, Nijmegen, the Netherlands
[c] University of Cambridge, Department of Psychology, Cambridge, United Kingdom
[d] Radboud University Medical Center, Department of Neurology, Nijmegen, the Netherlands
[e] Radboud University Medical Center, Department of Medical Psychology, Nijmegen, the Netherlands

ARTICLE INFO

Keywords:
Visual working memory
Binding
Computational modeling
Lesion-symptom mapping
Stroke

ABSTRACT

Visual memory for objects involves the integration, or binding, of individual features into a coherent representation. We used a novel approach to assess feature binding, using a delayed-reproduction task in combination with computational modeling and lesion analysis. We assessed stroke patients and neurotypical controls on a visual working memory task in which spatial arrays of colored disks were presented. After a brief delay, participants either had to report the color of one disk cued by its location or the location of one disk cued by its color. Our results demonstrate that, in the controls, report imprecision and swap errors (non-target reports) can be explained by a single source of variability. Stroke patients showed an overall decrease in memory precision for both color and location, with only limited evidence for deviations from the predicted relationship between report precision and swap errors. These deviations were primarily deficits in reporting items rather than selecting items based on the cue. Atlas-based lesion-symptom mapping showed that selection and reporting deficits, precision in reporting color, and precision in reporting location were associated with different lesion profiles. Deficits in binding are associated with lesions in the left somatosensory cortex, deficits in the precision of reporting color with bilateral fronto-parietal regions, and no anatomical substrates were identified for precision in reporting location. Our results converge with previous reports that working memory representations are widely distributed in the brain and can be found across sensory, parietal, temporal, and prefrontal cortices. Stroke patients demonstrate mostly subtle impairments in visual working memory, perhaps because representations from different areas in the brain can partly compensate for impaired encoding in lesioned areas. These findings contribute to understanding of the relation between memorizing features and their bound representations.

1. Introduction

Visual deficits are common following stroke and vary widely, from reduced acuity and visual field loss, to visual inattention and deficits in perceiving specific features (Beaudoin et al., 2013; Rowe et al., 2017). Visual perception is also the basis for visual working memory, the retention of visually perceived features and objects over a short period of time that is required for many everyday tasks and important for subsequent episodic memory formation. In this study, we investigated whether stroke affects visual working memory in cases where perception is spared. Moreover, we examined whether stroke can result in specific impairments of feature binding in working memory, that is, the ability to memorize which visual features belong to the same object.

Various previous studies have investigated the effects of ageing and

Abbreviations: fMRI, functional magnetic resonance imaging; LSM, lesion-symptom mapping; ROI, region of interest; BA, Brodmann area.

* Corresponding author at: Radboud University, Donders Institute for Brain, Cognition and Behaviour, Centre for Cognition (Neuropsychology & Rehabilitation Psychology), Montessorilaan 3, 6500 HE Nijmegen, the Netherlands.

E-mail addresses: S.Lugtmeijer@donders.ru.nl (S. Lugtmeijer), ss2361@cam.ac.uk (S. Schneegans), n.a.lammers@amsterdamumc.nl (N.A. Lammers), L.Geerligs@donders.ru.nl (L. Geerligs), FrankErik.deLeeuw@radboudumc.nl (F.E. de Leeuw), E.H.F.deHaan@uva.nl (E.H.F. de Haan), pmb20@cam.ac.uk (P.M. Bays), r.kessels@donders.ru.nl (R.P.C. Kessels).

[1] These authors contributed equally to this work.

https://doi.org/10.1016/j.nlm.2021.107387
Received 3 August 2020; Received in revised form 20 December 2020; Accepted 10 January 2021
Available online 15 January 2021

neurological conditions on feature binding in working memory. While some early work suggested that binding memory is selectively impaired in healthy ageing (Cowan, Naveh-Benjamin, Kilb, & Saults, 2006; Mitchell, Johnson, Raye, Mather, & D'Esposito, 2000), several more recent studies have consistently concluded that there is a general decline of working memory performance in older adults, but no specific impairment for feature-feature or feature-location binding (Brockmole, Parra, Della Sala, & Logie, 2008; Pertzov, Heider, Liang, & Husain, 2015; Rhodes, Parra, & Logie, 2016). In contrast, there is strong evidence that working memory for binding (object-location and object-color) compared to memory for individual features is selectively impaired in Alzheimer's disease, but not in other forms of dementia (Della Sala, Parra, Fabi, Luzzi, & Abrahams, 2012; Liang et al., 2016, Parra et al., 2015). Impairments in object-location binding have also been observed in patients with temporal lobe damage following a form of autoimmune encephalitis (Pertzov et al., 2013).

By studying feature binding in a stroke population we can get additional insight into brain regions necessary for feature recall and binding in visual working memory. Unlike Alzheimer's disease and encephalitis, stroke has a sudden impact on the brain with lesions that can be identified with respect to site and size, which makes the etiology more suited to make inferences on neural correlates of deficits (De Haan & Karnath, 2018). Stroke may result in focal lesions on the one hand, but on the other hand also in wide-spread disruptions of network activation (Adhikari et al., 2017). Consequently, we expected to find specific impairments in visual working memory as a consequence to a lesion in a key region, in addition to stroke patients as a group differing from stroke-free controls. Lesion-behavior mapping was used to identify these regions.

In the present study, we assess visual working memory performance for color, location, and the binding of colors to locations (a form of extrinsic intra-item binding; cf. Piekema, Rijpkema, Fernández, & Kessels, 2010) in stroke patients and age-matched controls. We employ a novel approach that combines behavioral testing in two delayed reproduction tasks with computational modeling. This approach makes use of recent findings on feature binding in healthy participants to measure both memory precision for individual features and to detect specific impairments in feature binding.

In the delayed reproduction task, participants briefly view a visual array of sample stimuli, and after a short delay interval have to report a feature of a cued item on a continuous scale (e.g. by adjusting the color of a probe to match the sample item at a cued location; Wilken & Ma, 2004). This type of task allows us to distinguish different types of recall errors. The dispersal of reported features around the true value of the cued target item yields a sensitive measure of memory precision for individual features. Failures to retrieve the binding between features are typically associated with swap errors, in which participants report the feature of a sample item other than the target (Bays, Catalao, & Husain, 2009). Importantly, however, an elevated proportion of swap errors does not necessarily demonstrate a specific deficit in binding memory, but may also reflect an impairment in memory for the cue features that are used to indicate which item is the target (Pertzov et al., 2015).

Indeed, several studies have found that swap errors occur more frequently if memory for the cue feature is less precise (e.g. using color cues instead of location cues; Rajsic & Wilson, 2014), and that swap errors are more likely across items that are similar in their cue feature (Bays, 2016; Emrich & Ferber, 2012; Rerko, Oberauer, & Lin, 2014). Two recent computational models have incorporated imprecision in memory for the cue feature as a key cause of swap errors, and have successfully accounted for the specific patterns of swap errors in various delayed reproduction tasks (Oberauer & Lin, 2017; Schneegans & Bays, 2017).

Here, we employ the neural binding model of Schneegans and Bays (2017) as an analytical tool to assess memory for feature binding. It is based on a population coding account that has recently been shown to perform equal or better in fitting behavioral data than competing models of visual working memory (Schneegans, Taylor, & Bays, 2020) while also being firmly grounded in neural principles. The model proposes that each item's cue and report features are jointly represented in a neural population code, and it explains recall errors as a result of decoding from noisy neural activity. Imprecision in recall arises from variability in decoding the reported feature, and swap errors arise from decoding variability in the cue feature.

One key property of the neural binding model is that it makes predictions across task conditions. If the roles of cue and report feature are swapped, the same decoding variability that was previously causing recall imprecision should now be reflected in the frequency of swap errors, and vice versa. In the present study, participants either had to report the color of an item cued by its location, or its location when cued with its color. We validated the predictions of the neural binding model in the control group, and then used model fits to detect deviations from the predicted relationship between recall precision and swap errors in the patient group. Such deviations served as a measure of specific impairments either in binding or feature recall.

2. Materials and methods

2.1. Participants

Eighty-eight patients from the Functional Architecture of the Brain for Vision (FAB4V) study, a multi-centre prospective cohort study on vision and cognition after ischemic stroke in adults, participated in this study. The Medical Review Ethics Committee Utrecht approved the study (30-06-2015), and written informed consent according to the Declaration of Helsinki, was obtained from all participants prior to participation. Exclusion criteria were: haemorrhagic stroke, cerebral venous sinus thrombosis, pre-existing cognitive decline (score $\geq$ 3.6 on the Dutch version of the Informant Questionnaire on Cognitive Decline in the Elderly [IQCODE], filled in based on functioning before the stroke; Schmand, Ooms, & Ribbe, 1997) or dementia, psychiatric disorder, severe aphasia, pre-existing visual impairment, and disrupted perception as a consequence of stroke, like hemianopsia. Cognitive measurements took place between April 2016 and March 2020. Patients were at least 4 weeks post-stroke, 46 patients were assessed on visual working memory in the subacute stage within 6 months, 42 in the chronic stage between 6 months and 3 years after stroke (range in days 29–1,055, median 106 days). All patients underwent an MRI scan within 6 months post-stroke (range in days 17–186, median 52 days). Based on the Bells test (Gauthier, Dehaut, & Joanette, 1989) none of the patients had neglect. Five patients were excluded from analyses because of missing data due to technical or logistic reasons ($N = 3$) and fatigue ($N = 2$), resulting in a total sample size of 83 patients (*M*age (*SD*) = 63 (11); m:f = 61:22; *M*education [range] = 5.3 [2–7]).

A stroke-free control group ($N = 88$), matched for age, was recruited, without a history of neurological disease or cognitive decline (self-report). Thirteen controls were excluded because of incomplete data (all due to technical or logistic reasons) resulting in a control group of 75 subjects (*M*age (*SD*) = 60 (11); m:f = 38:37; *M*education [range] = 5.9 [4–7]). Controls were recruited via social networks or were spouses or family members of patients.

All participants had normal or corrected-to-normal visual acuity, and none reported color blindness.

2.2. Paradigm

Participants performed a delayed reproduction task that assesses memory precision for color and location and for binding between those features (adapted from Schneegans & Bays, 2017, experiment 1). In one condition a location cue was given and participants were instructed to report the corresponding color, in the other condition a color cue was given and the location needed to be reported (Fig. 1). At the beginning of each trial a white fixation cross (diameter 0.75° of visual angle) was

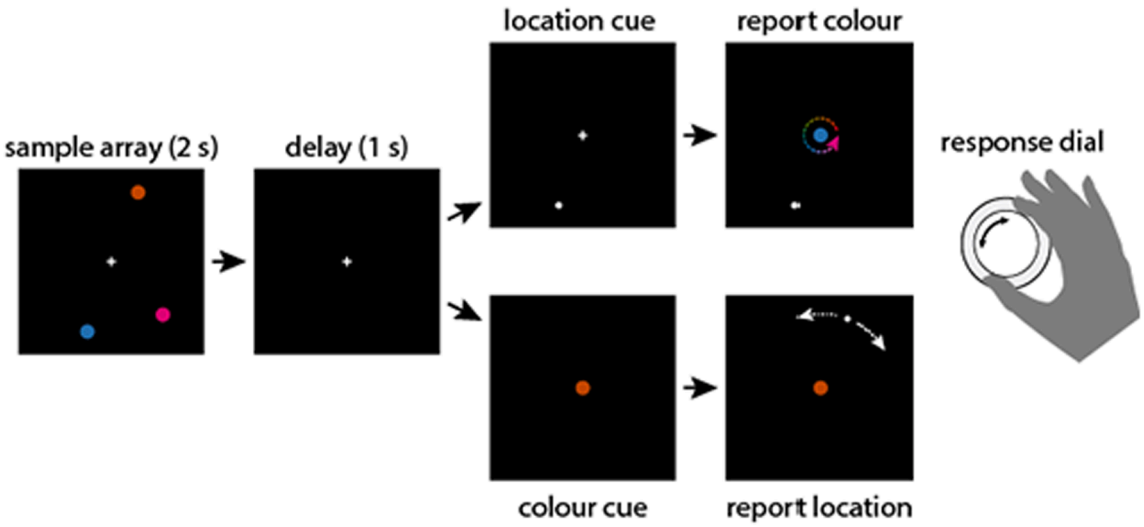


Fig. 1. Delayed reproduction task with two conditions. The sample array consists of three disks with randomly chosen colors and locations on an imaginary circle around the fixation point. In the report-color condition a location is cued and participants use a response dial to report the matching color. In the report-location condition a color cue is given and participants use the dial to adjust a small white disk to the matching location.

presented for 2 s. This was followed by a sample array showing three colored discs (0.5° radius) positioned on an imaginary circle (6° radius), cantered on the fixation cross. Locations were randomly selected for every trial with a minimum distance of 30° to neighbors. Colors were chosen at random from a color wheel, defined as a circle in Commission Internationale de l'Eclairage (CIE) L*a*b* coordinates with constant luminance (L* = 50), center at a* = b* = 20, and radius 60. Colors for the different discs were separated at least 30° on the color wheel. After a presentation duration of 2 s, a black display with fixation cross was presented for 1 s followed by a cue. One of the three discs from the sample array was selected as the target. In the *report-color* condition, a white disc (0.25° radius) appeared at the location of the target item. Participants adjusted the color of a centrally presented disc (0.75° radius) by cycling through the color wheel until the color matched the recalled color of the target. In the *report-location* condition, the cue was a centrally presented disc that matched the color of the target. Participants adjusted the location of a white disc (0.25° radius) on the imaginary circle to match the target's recalled location. Participants responded using an input dial (PowerMate USB Multimedia Controller, Griffin Technology) and were instructed to answer precisely rather than fast. All participants were capable of operating the input dial without aid. Participant performed one block of 40 trials per task condition, each preceded by six single item practice trials. The order of the blocks was counterbalanced across participants.

2.3. Behavioral data processing

Stimulus locations and colors were analyzed and are reported with respect to the circular feature space of possible values, -π to π radians. Recall error was calculated as the distance in radians between the reported value and the true feature value of the target item. Deviation between the response and feature values of non-target items in each trial was calculated to assess evidence for swap errors (erroneous report of the feature of a non-target item). Histogram plots of non-target deviations were corrected for the effects of minimum feature distance between items within a trial (Schneegans & Bays, 2017; see S1 for details).

2.4. Neural binding model

The neural binding model (Schneegans & Bays, 2017) extends a neural population model of working memory (Bays, 2014) to explain patterns of swap errors in delayed reproduction tasks. It assumes that the features of all sample items are encoded in the spiking activity of an idealized population of neurons with conjunctive coding (Fig. 2A). Each neuron's mean activity is determined by its tuning functions for both stimulus color and stimulus location, modelled as von Mises distributions with different concentrations for the two features.

Spiking activity for each neuron is generated by an independent Poisson process.

When a cue is presented in the delayed reproduction task, the feature values of all sample items are decoded from the noisy spiking activity by maximum likelihood estimation. The item whose decoded cue feature value is closest to the given cue is selected, and its decoded report feature value is generated as a response. Variability in decoding leads both to imprecision in the reported value and to swap errors (Fig. 2B), which occur when the decoded cue feature value of a non-target item is closer to the given cue than the decoded cue feature value of the actual target (see S2 for full model description and derivation of response distributions).

The model has three free parameters. The widths of the tuning curves for color and for location determine the decoding precision in each feature dimension (with sharper tuning curves leading to lower decoding errors, as illustrated in Fig. 2A). The gain parameter determines the mean spike rate, which globally scales decoding precision and controls variability of precision across trials.

We obtained separate model fits for each participant's data in the two task conditions, as well as a combined fit across task conditions for each participant. In the latter fit, the same neural population underlies both color and location report, the only difference being which feature is used as cue and which is reported. This model makes predictions about the relationship between error distributions in the two tasks (Fig. 2C, left column). Concretely, the distribution of decoding errors for color (reflected by the red solid line in Fig. 2A) directly matches the distribution of response errors around the selected item in the report-color condition (width of the peak in Fig. 2C, top left). The same distribution of decoding errors determines the likelihood of a specific non-target item being selected for response generation in the report-location condition, dependent on how similar that non-target is to the target in color (Fig. 2B). A broader distribution of decoding errors leads to more swap errors, which are reflected by a central peak in the distribution of response deviations from non-targets (Fig. 2C, bottom left). The same is true analogously for location, which typically shows a narrower distribution of response errors and fewer swap errors. The combined fit of the

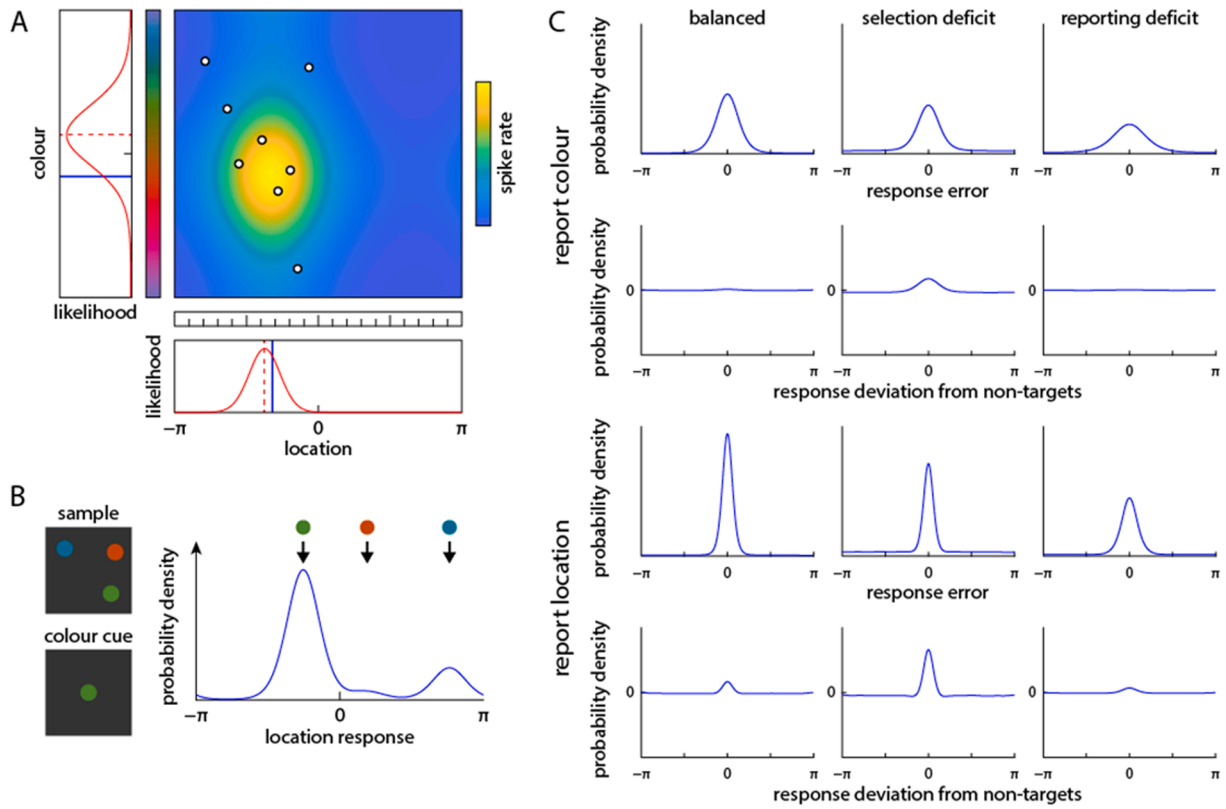


Fig. 2. Neural binding model. (A) The features of a sample stimulus (blue lines in plots for color and location) are encoded in the spiking activity of a neural population with conjunctive coding. The spike rate of each neuron is shown color coded as a function of the neuron's preferred values for color and location, and white dots indicate spikes occurring in an example trial. Maximum likelihood decoding from the spiking activity yields noisy estimates of the true values (red dashed lines). The width of the likelihood distributions (red solid lines) provides an estimate of decoding precision for each feature. (B) The model predicts a response probability distribution for each trial, with decoding variability in the report feature leading to imprecision in responses, and decoding variability in the cue feature dimension leading to swap errors depending on cue similarity. (C) Model predictions for distributions of response errors are obtained by averaging response probability distributions over many trials, aligned to the target or non-target report feature values. Effects of the binding index are shown for a model with other parameters held fixed at typical values for this task. Left column, the model with no deficit shows broad response error distribution with almost no swap errors in the color report, and a sharper distribution with some swap errors (indicated by the central peak in the distribution of response deviations from non-targets) in the location report. A selection deficit (middle column) increases the proportion of swap errors in both feature dimensions, which leads to longer tails in the response error distribution without affecting the shape of the central peak. A reporting deficit (right column) produces broader distribution of response errors, but does not affect the proportion of swap errors.

model to both task conditions takes all forms of errors into account to estimate the decoding precision for color and location that best explain a participant's response errors in each single trial.

2.5. Extended neural binding model

We extended the neural binding model to detect specific deviations from the predicted relationship between report precision and swap errors. We allow for the possibility that only a certain proportion of all available spikes contribute to the selection of the presumed target item (with others only contributing to the decoding of the report feature value), leading to a selective impairment in selecting the target item based on the given cue. We also allow for the converse effect, in which all spikes contribute to the selection of the cued item, but only a subset can be used in decoding the report feature value, resulting in a selective decrease in report precision.

These adjustments are parameterized with a new free parameter in the model, the *binding index* (Fig. 2C). With a binding index of zero, the model behaves exactly like the original neural binding model, with the proportion of swap errors in one task condition matched to the report precision in the other condition (left column in Fig. 2C). A negative binding index signifies a selection deficit, in which fewer spikes contribute to the decoding of the cue feature values (middle column in Fig. 2C). This yields an increased proportion of swap errors (larger peaks in the distribution of response deviations from non-targets), while the report precision remains the same (the width of response error distributions is unchanged, although long tails appear due to the occurrence of swap errors). An index of −1 indicates maximum selection deficit, for which the selection of the cued item is completely random. A positive binding index signifies a reporting deficit, in which fewer spikes contribute to the decoding of the report feature (right column in Fig. 2C). This yields broader response error distributions without increasing the frequency of swap errors. A value of 1 indicates maximum reporting deficit, for which the response distributions are always uniform.

In addition to this binding index, we derive measures of memory precision for color and location from the model fits of each participant. We use the circular standard deviation of the decoding error for each feature as precision measure, computed from the fitted model but excluding any selection or reporting deficits (i.e., computed as if the binding index was zero). This means that for participants with a selection deficit in the model fits, the precision reflects how closely their responses are clustered around the items' true feature values when reporting that feature (ignoring swap errors). For participants with a reporting deficit, the precision reflects how precise their memory for a feature has to be to explain their reliability in selecting the correct item based on the given cue.

2.6. Statistical analysis

Hypothesis testing was conducted using t-tests. Models were compared using the Akaike information criterion with correction for small sample size (AICc), computed for each participant based on the maximum likelihood fit of each model (see S1 for details).

2.7. Neuroimaging

Participants underwent a 3-T MRI scan, at the Radboudumc and UMCG on the Siemens Magnetom Prisma, at the Amsterdam UMC and UMCU on the Philips R5. For the Siemens scanners, sequence details were as follows: T2 FLAIR (TI = 1650 ms, TR = 4800 ms, TE = 484 ms, [FOV] = 280 mm, voxel size 0.9 × 0.9 × 0.9 mm^3). For the Philips scanners, sequence details were: T2 FLAIR (TI = 1650 ms, TR = 4800 ms, TE = 253 ms, [FOV] = 250 mm, voxel size 1.12 × 1.12 × 0.56 mm^3).

Lesions were manually delineated using ITK-SNAP software (Yushkevich et al., 2006). The FLAIR and binary lesion mask were normalized to an older adult MNI template using the plug-in clinical toolbox for SPM (Crinion et al., 2007; Rorden, Bonilha, Fridriksson, Bender, & Karnath, 2012).

2.8. Lesion analyses

For the associations between lesion location and outcome measures we used the three performance measures obtained from the extended neural binding model fits: binding index, memory precision for color, and memory precision for location (circular standard deviation of the decoding error).

Atlas-based LSM (lesion-symptom mapping) was used to investigate which lesions are associated with a reporting or selecting deficit, and with memory precision. For the binding index the association with lesion location was tested two-sided. For memory precision, associations were tested one-sided, with higher behavioral scores indicating worse performance. Statistical lesion analysis software NiiStat was used (https://github.com/neurolabusc/NiiStat). Atlas-based analysis is based on the cumulative lesion burden in a specific ROI (region of interest), instead of investigating lesions on a voxel-wise basis. The advantage is effectively increasing the number of areas that have sufficient coverage across participants. The assumption is that lesions in the same ROI affect behavior in the same way. In addition, a ROI-based approach reduces the strict control needed for multiple testing that is required in voxel-wise analyses. Cortical ROIs are based on Brodmann areas (BA). The Brodmann atlas has relatively large ROIs compared to other commonly used atlases. This has the advantage of making maximum use of cumulative lesion burden to cover a large part of the cerebral cortex. For white-matter ROIs the CAT atlas was used as this is the most widely used white-matter atlas (Catani & De Schotten, 2008; https://www.natbrainlab.co.uk/).

Only ROIs with a lesion coverage of at least 4 (6%) subjects were included. To correct for multiple comparisons permutation testing was set to 5,000 permutations at $p < .025$ for the binding index (two-sided) and at $p < .05$ for reporting precision (one-sided). Lesion volume, age, education, gender, interval between stroke and MRI, and interval between MRI and assessment were included as covariates and were regressed on both behavioral and lesion data (DeMarco & Turkeltaub, 2018).

2.9. Data availability

Anonymized data are available upon request a year after the completion of the project (01-07-2021).

3. Results

To investigate visual working memory after stroke, we assessed stroke patients and age-matched controls in two delayed-reproduction tasks. Participants viewed a sample array of colored disks, and after a brief delay either had to report the color of one disk cued by its location (report-color condition) or the location of one disk cued by its color (report-location condition). We fitted a neural population model to the behavioral data to detect specific deficits either in memory for feature bindings or in the ability to report memorized features. Second, we investigated which lesion locations were associated with recall precision and with deficits in reporting or selecting features.

3.1. Behavioral data

The behavioral data are shown in Fig. 3. Consistent with previous studies, the distribution of response errors is broader overall in the report-color (Fig. 3A) than in the report-location condition (Fig. 3D). Nonetheless, the error distribution for location reports shows long tails, with above-zero proportion of responses even at the largest deviations from the target location. This is consistent with the presence of swap errors, in which participants incorrectly report the location of a sample item that is not the cued target.

The occurrence of such swap errors can be detected by plotting the histogram of response deviations from the report feature values of all non-target items in each trial (Fig. 3B and E). In the absence of swap errors, these distributions should be uniform (after correcting for effects of minimum distance between items' feature values). However, in the report-location condition (Fig. 3E) we observed a significant central tendency that indicates a clustering of responses around the locations of non-target items and thus the occurrence of swap errors (mean absolute deviation from non-target features compared to value expected by chance if there were no swap errors: patients *M (SD)* = 1.74 (0.10) vs chance 1.78 (0.07), $t(82) = 5.82$, $p < .001$, controls 1.77 (0.10) vs chance 1.80 (0.07), $t(74) = 4.05$, $p < .001$; see S1 for method details). In the report-color condition the central peak is absent, and we instead find a small, but significant tendency to avoid the colors of non-target items (Fig. 3B; mean absolute deviation: patients 1.77 (0.12) vs chance 1.75 (0.08), $t(82) = 3.18$, $p = .002$, controls 1.79 (0.09) vs chance 1.77 (0.07), $t(74) = 2.99$, $p = .004$).

Despite the overall similarity in response distributions between the two participant groups, we find that recall performance is impaired in stroke patients (Fig. 3C and F). An independent-sample *t*-test shows that the circular standard deviation as a measure of variability is higher (indicating lower precision) in both conditions for patients compared to controls (report-color: patients *M (SD)* = 0.76 (0.35); controls *M (SD)* = 0.58 (0.28); $t(156) = 3.60$, $p < .001$; report-location: patients *M (SD)* = 0.56 (0.25); controls *M (SD)* = 0.47 (0.21); $t(156) = 2.45$, $p = .02$).

3.2. Model fits

We fitted the single-trial data of participants with a neural population model that has previously proved successful in capturing performance on similar tasks (Schneegans & Bays, 2017). The model assumes that the location and color of each item are encoded together in a conjunctive population code, such that each spike from this population yields a sample of both features of an item. Variability in decoding the memorized feature values from noisy spiking activity is used to explain both imprecision in reporting a target feature and the occurrence of swap errors, which is due to uncertainty in selecting an item based on the given cue.

We found that, for the majority of participants, performance in both conditions is well explained by a single conjunctive population code, varying only in which feature is used as a cue and which is to be reported. This combined model provided a better quality of fit than separate neural binding models fitted to each task condition independently, as measured by AICc scores (patients: combined better than separate fit for 68 out of 83, mean ΔAICc = 1.71; controls: 67 out of 75, mean ΔAICc = 2.82). This supports the hypothesis that a single source of

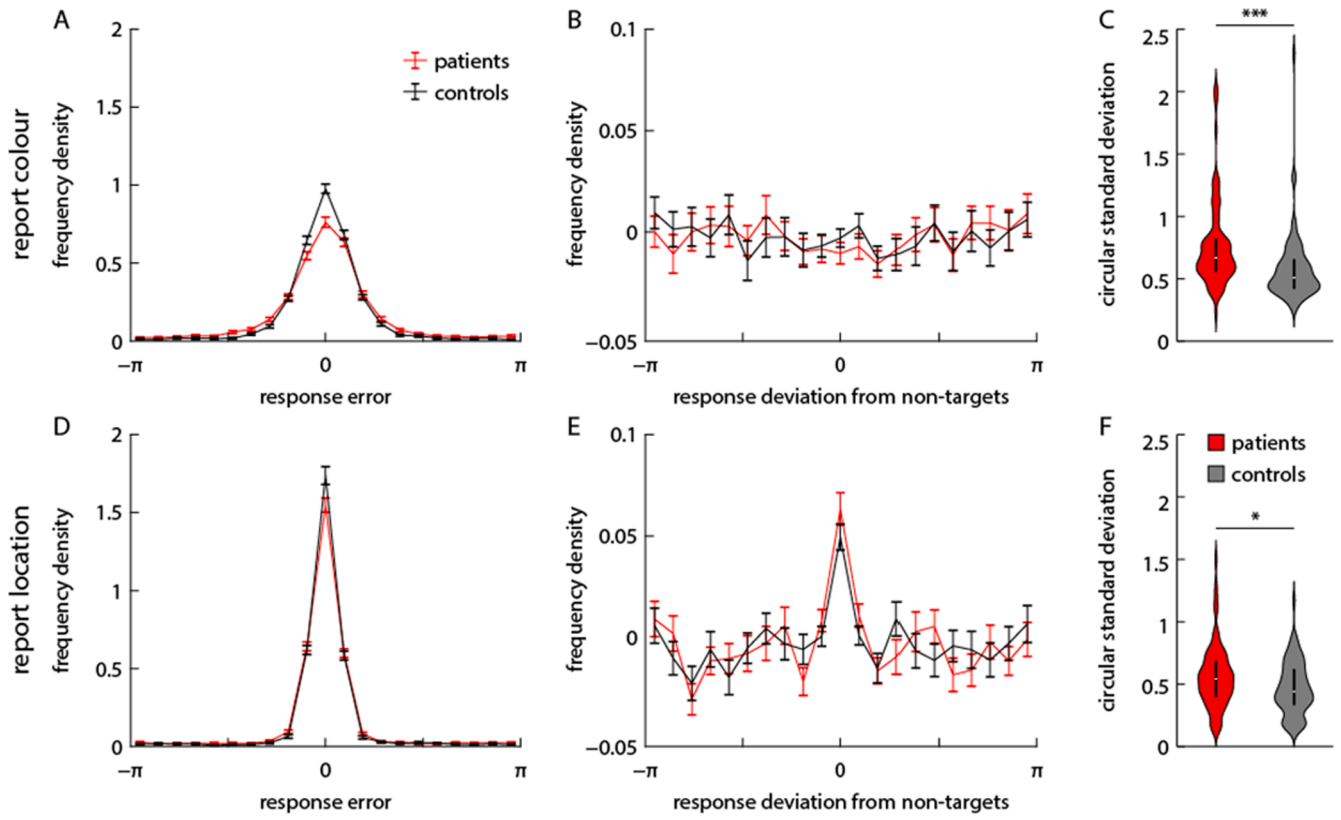


Fig. 3. Behavioral data. Data for the report-color condition (A-C) and the report-location condition (D-F). From left to right: histogram of response errors, histogram of response deviations from report feature values of non-targets in each trial, and distribution of response precision (as circular standard deviation, higher values indicating less precise) in each participant group. Violin plots show the median (white dot), interquartile range (black line), and kernel density plot (Gaussian kernel with bandwidth of 0.05).

recall variability for each feature can explain both report imprecision and swap errors in the two task conditions.

Because stroke is a heterogeneous syndrome, we wanted to be able to quantify specific memory deficits in individual stroke patients. We therefore relaxed the neural binding model's assumption that there is a fixed relationship between report precision in one task condition and frequency of swap errors in the other condition. We introduce a binding index as a new parameter that can capture specific impairments either in selecting a cued item or in reporting the feature of an item once it is selected, by controlling what proportion of spikes contributes to item selection and report feature decoding, respectively. This binding index can take values between −1 (maximum selection deficit) and 1 (maximum reporting deficit).

Fig. 4 shows fits of both the original neural binding model (blue) and the extended model (red) to the response distributions of selected patients illustrating different forms of recall deficits. Model fits for participants in Fig. 4A and B produce values of the binding index close to zero, indicating neither selection nor reporting deficits, and consequently show near identical fits of the two models. Note that these participants still differ substantially in their memory precision for both color and location, but in both cases the frequency of swap errors is consistent with the report precisions. In contrast, the individuals in Fig. 4C and D show specific selection deficits, which suggests that memory for the binding between cue and report feature is impaired: the frequency of swap errors is higher than would be expected based on the participant's reporting precision in each feature. The original binding model in these cases fails to fully capture the observed proportion of swap errors, since it is constrained by the reporting precision. Finally, participants in Fig. 4E and F show evidence for specific reporting deficits. Both of these patients were able to use a color cue to reliably select the target item for the location report (indicating that they held the colors and their binding to locations in memory), but performed very poorly when reporting colors. In these cases, the model without binding index is forced to overestimates the frequency of swap errors in the report-location condition in order to better capture the very broad error distribution in the color report. (In the case of Fig. 4F, the extended model still overestimates the proportion of spatial swaps to a lesser degree, since it is still constrained by the assumption that reporting deficits are symmetrical between the two conditions.)

Model comparison using AICc scores showed that a non-zero binding index was preferred only for a small number of participants (patients: 12 out of 83, mean ΔAICc = 1.16 in favor of the original model; controls: 7 out of 75, mean ΔAICc = 1.41 in favor of the original model). This is expected if only a few participants show selective impairments of binding in visual working memory.

Fig. 5 shows the distribution of performance measures derived from the model fits for patients and controls, namely estimates of memory precision for each feature dimension and binding index. Independent sample t-tests show that patients have a significantly lower memory precision for both color and location compared to controls (color: patients *M (SD)* = 0.57 (0.26); controls *M (SD)* = 0.46 (0.15); $t(156) = 3.28$, $p = .0013$; location: patients *M (SD)* = 0.30 (0.09); controls *M (SD)* = 0.27 (0.08); $t(156) = 2.18$, $p = .031$).

The difference in binding index between patients and controls did not reach significance at the group level (patients *M (SD)* = 0.11 (0.30); controls *M (SD)* = 0.02 (0.27)); $t(156) = -1.80$, $p = .07$). Based on visual inspection of the results, we tested post hoc whether group means deviate from 0. The estimates of the binding index were significantly shifted towards the positive range in patients, indicating an overall tendency towards reporting deficits (single sample *t*-test, $t(82) = 3.25$, $p = 0.002$). In contrast, estimated binding indices were not significantly different from zero in controls ($t(74) = 0.79$, $p = 0.44$). Moreover, if we

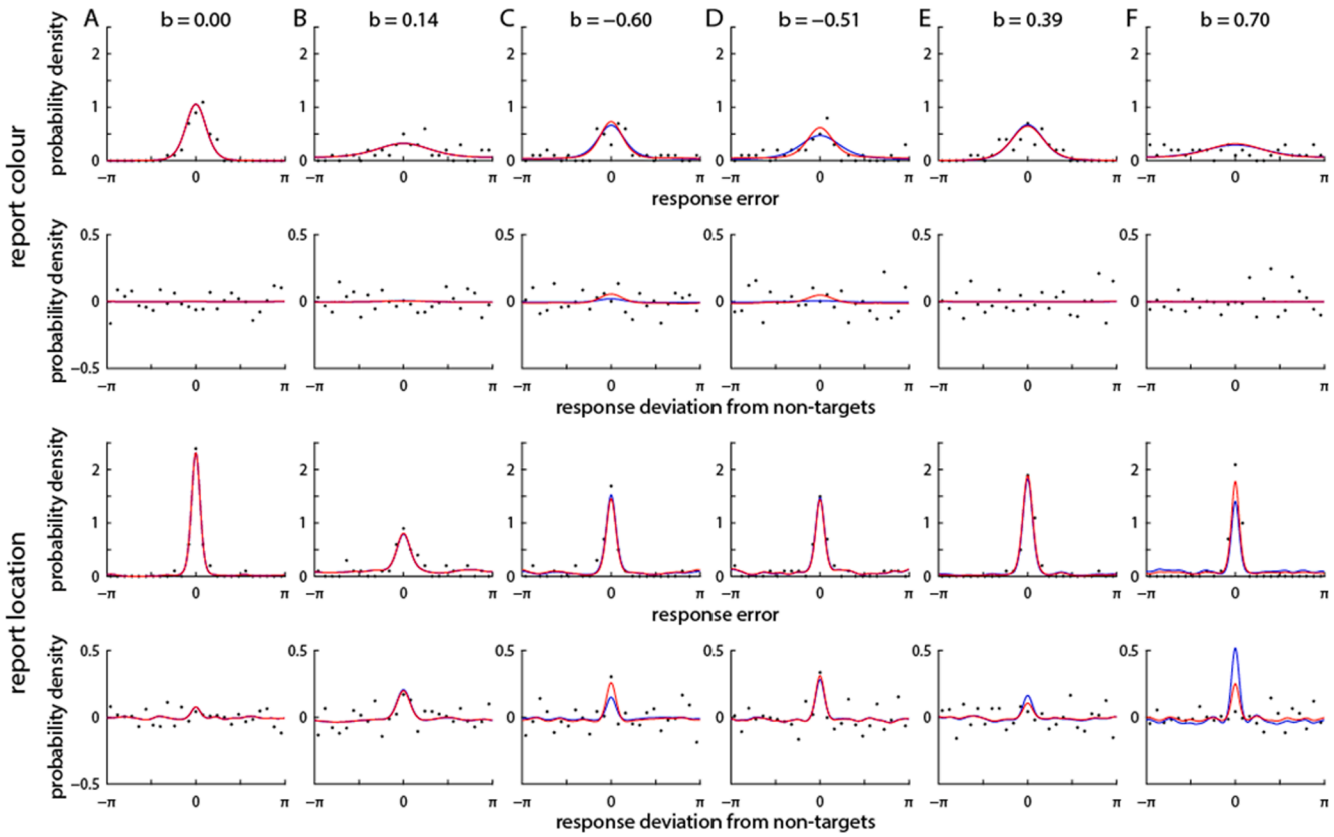


Fig. 4. Fits of neural binding model to behavioral data of individual patients. Response histograms of patients are shown as black dots, fits of the original neural binding model in blue and fits of the extended model in red. (A, B) Patients without selection or reporting deficit. The binding index is close to 0 so the model fits are near-identical. (C, D) Patients with selection deficit. The binding index is negative and the extended model fit indicates more binding errors than the original model fit. (E, F) Patients with reporting deficit. The binding index is positive and the extended model fit indicates less binding errors than the original model fit.

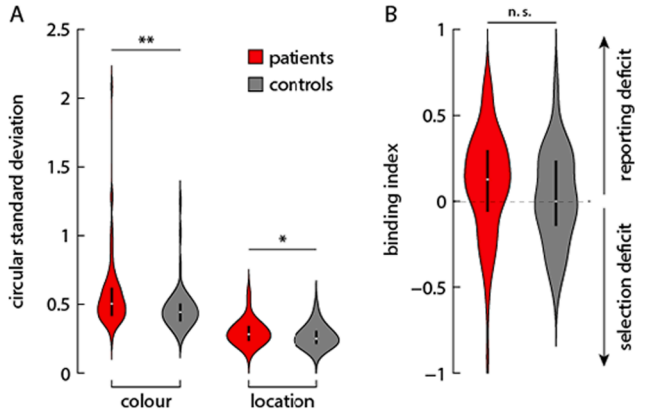


Fig. 5. Model based performance measures. Measures are based on maximum likelihood fits of the neural binding model to behavioral data of patients and controls. (A) Memory precision in each feature dimension. (B) Binding index, with positive values indicating a reporting deficit and negative values a selection deficit. Violin plots as in Fig. 3, with bandwidth of 0.05 for circular standard deviation and 0.1 for binding index.

consider only participants in which the introduction of the binding index improves quality of fits, a majority showed a reporting deficit (positive binding index for 10 out of 12 patients and 4 out of 7 controls), indicating that specific deficits in feature selection are relatively rare.

We note that despite the overall tendency towards a reporting deficit, patients still showed a higher estimated proportion of swap errors when reporting location (patients *M (SD)* = 0.05 (0.05); controls 0.04 (0.04); $t(156) = 2.76$, $p = .04$), due to the overall lower memory precision. In the color report, the difference in swap frequency was not significant (patients *M (SD)* = 0.02 (0.04); controls *M (SD)* = 0.01 (0.02); $t(156) = 3.73$, $p = .09$).

3.3. Lesion analyses

Of the total sample of 83 patients 65 were included in these analyses. Eight were excluded because of a missing structural MRI scan, seven had no clear lesion on the MRI scan, three patients had widespread white matter hyperintensities. Median lesion volume was 3.06 cm3 (range 0.02–85.12 cm3). Lesions in the left hemisphere were most common ($N = 28$), followed by lesions in the right hemisphere ($N = 21$), bilateral lesions ($N = 13$) and brain stem lesions ($N = 3$). Fig. 6 shows the lesion prevalence map.

3.3.1. Atlas-based lesion-symptom mapping

Of 82 cortical ROIs included in the Brodmann atlas, 37 (22 right hemisphere) were covered by at least 4 lesions. Twenty-seven white-matter tracts of 32 defined in the CAT atlas had sufficient lesion coverage (see S3 for details). Behavioral variables of interest were binding index, and precision of reporting color and reporting location indicated by the circular standard deviation. Lesion volume, age, education, gender, interval between stroke and MRI, and interval between MRI and assessment, were included as covariates. Covariates were corrected for by regressing them on the behavioral and lesion data.

Lesion status in BA3, the primary somatosensory cortex, in left hemisphere was negatively associated with the binding index, indicating selection deficits (threshold $z < -3.00$ and $z > 2.66$, $z = -3.12$). Precision in the report-color condition was associated with BA6, premotor and supplementary motor cortex, in the right hemisphere (threshold $z >$

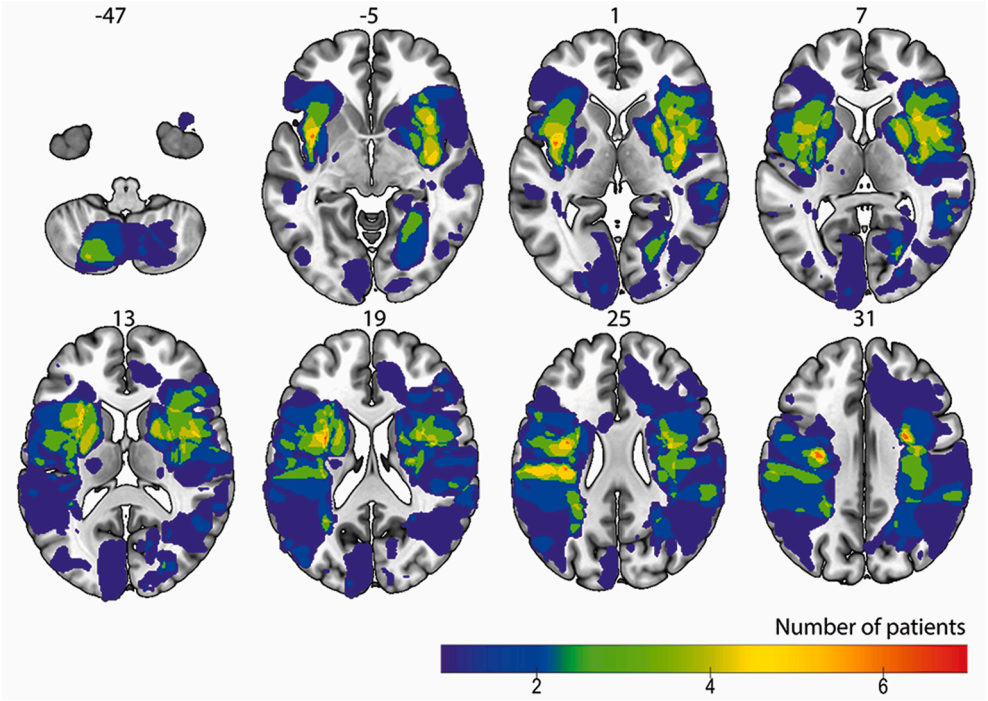


Fig. 6. Lesion density plot. Maximum overlap 7. MNI coordinates are specified for each axial slice. Left hemisphere is depicted on the left.

4.57, $z = 4.62$), and with BA44, Broca's area, also in the right hemisphere (threshold $z > 4.57$, $z = 5.24$). In the left hemisphere precision in reporting color was associated with BA7, superior parietal lobe, in the left hemisphere (threshold $z > 4.57$, $z = 5.64$) and BA41, auditory cortex (threshold $z > 4.57$, $z = 5.69$), the posterior segment of the arcuate fasciculus (threshold $z > 4.33$, $z = 5.88$), and the optic radiation (threshold $z > 4.33$, $z = 5.04$). For precision in the report-location condition there were no significant neural correlates. Results are summarized in Table 1 and Fig. 7.

4. Discussion

Mechanisms underlying the binding of visual features have been studied in cognitively unimpaired individuals, as well as in ageing and patient populations (for a review see Schneegans & Bays, 2019). In the current study we investigated feature recall and binding in visual working memory in stroke patients to assess if specific deficits in visual working memory in this population occur and if so, what the neuroanatomical basis is. Specifically, we tested performance in two tasks requiring binding of stimulus colors to locations. This is typically classified as a form of relational binding, in contrast to conjunctive binding which refers to the binding of different visual features within an object. The distinction is partly motivated by findings from long-term memory research, where distinct forms of binding deficits have been associated with different clinical conditions (Moses & Ryan, 2006; Mayes, Montaldi, & Migo, 2007). In working memory research, selective impairments of conjunctive binding have only been described for patients suffering from Alzheimer's disease (Della Sala et al., 2012). In contrast, impairments in relational binding have been described for different clinical conditions when testing binding of objects to location (Olson, Page, Moore, Chatterjee, & Verfaellie, 2006; Pertzov et al., 2013, Liang et al., 2016), and in a study testing binding between separately presented colors and objects in a single patient with a stroke affecting the hippocampus (Parra et al., 2015).

Based on these earlier results, we expected that stroke damage in different cortical regions may result in selective impairment in color-location binding. While we found that stroke patients showed an overall decrease in memory precision for both color and location, binding deficits – operationalized as a specific impairment in selecting the correct sample item based on the given cue – were rare. Visual deficits were associated with a distributed network of brain regions.

We used a novel approach to assess deficits in feature binding, using two symmetrical conditions of a delayed reproduction task in combination with computational modeling and lesion analysis. We built on theoretical work and behavioral results in healthy adults suggesting that the same variability in memory retrieval can account for errors in reporting a feature and for swap errors when the feature is used as a cue (Bays et al., 2009; Oberauer & Lin, 2017; Schneegans & Bays, 2017). Previous studies have qualitatively controlled for the effect of memory imprecision in the cue feature (e.g., by having participants first select a shape that was present in the sample array before reporting its location; Pertzov et al., 2015; Pertzov et al., 2013). However, the design used here enables quantification of memory precision for a visual feature when it has to be reported and when it is used as cue, in order to determine whether an additional source of binding errors is needed to explain the

Table 1
Results from the atlas-based lesion symptom mapping analysis.

ROI (Fig. 7)	Description	*N* subjects	Outcome measures
BA3 left (1)	Primary somatosensory cortex/ postcentral gyrus	7	binding index
BA6 right (2)	Premotor and supplementary motor cortex	11	precision color
BA7 left (3)	Visuo-motor coordination/ superior parietal lobe	4	precision color
BA41 left (4)	auditory cortex	5	precision color
BA44 right (5)	Broca's area/inferior frontal gyrus	9	precision color
Posterior segment arcuate left (6)	Connecting the inferior parietal lobe to Wernicke's area	6	precision color
Optic radiation left (7)	Connecting the lateral geniculate nucleus to the primary visual cortex	13	precision color

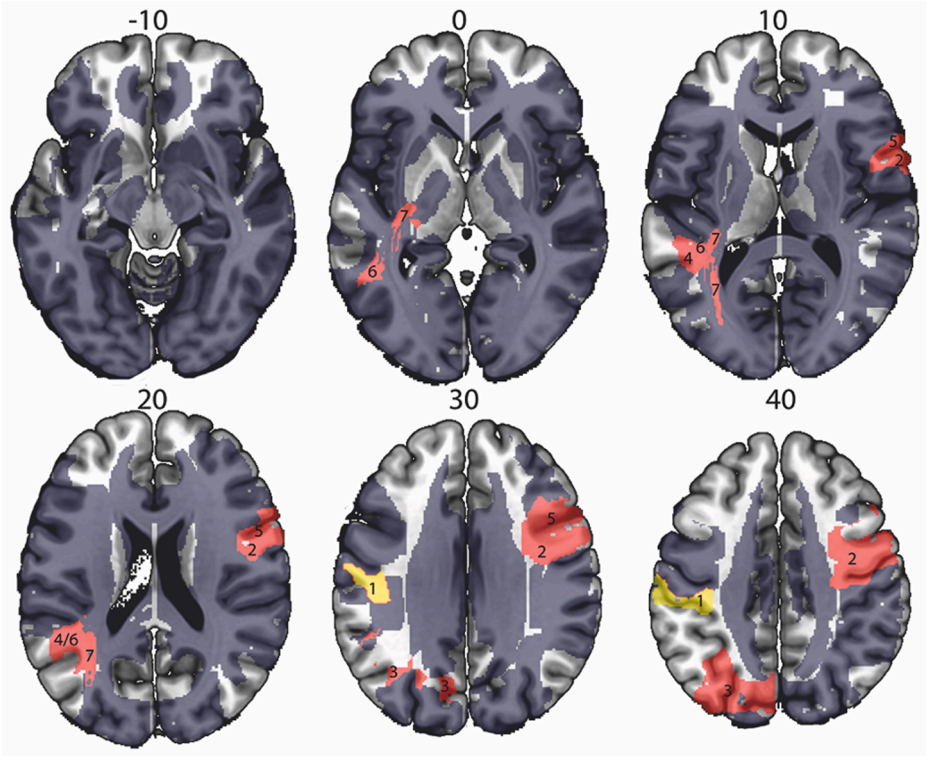


Fig. 7. Results of the atlas-based LSM analysis. Shaded areas show which regions are included in the analysis (≥4 overlapping lesions). ROIs significantly associated with the binding index are plotted in yellow; ROIs significantly associated with precision in reporting color in red. No areas were significantly associated with precision in reporting location. The numbers correspond to Table 1. Coordinates correspond to MNI standardized space. Left hemisphere is depicted on the left.

behavioral results.

We fitted the behavioral data with an existing neural population model which assumes that a single memory representation based on a conjunctive population code supports binding and recall. We also extended this model by introducing a binding index that can capture and quantify specific deviations from the assumption of a single source of variability. The implementation of such deficits in the binding model is broadly compatible with models assuming mixed representations with only a subset of neurons contributing binding information (Bouchacourt & Buschman, 2019; Matthey, Bays, & Dayan, 2015). However, given that lesion effects vary substantially between participants, we refrained here from trying to specify a concrete neural mechanism for observed impairments, and confined ourselves to quantifying the memory deficits as deviations from the existing model.

The behavioral results for the control group confirmed the assumption made in previous work that report imprecision and swap errors are explained by a single source of variability in healthy individuals. This was supported both by model comparison between the original and the extended binding model and parameter estimates for the binding index in the extended model. In patients, we found an overall decrease in memory precision for both color and location, but only limited evidence for deviations from the predicted relationship between report precision and swap errors. Critically, the deviations that we observed tended to be in the direction of a reporting deficit—patients tended to be worse at explicitly reporting a memorized feature value (especially for colors) than at using it as a cue to retrieve another feature of a matching item. This shows that selective impairments in memory for feature binding are rare for the patient group analyzed here.

One could argue that the observed reporting deficits in patients are due to the response procedure, in particular in the report-color condition where participants had to adjust the color of a central probe stimulus via a response dial. This might induce stronger interference than e.g. selection from a color wheel with all response options visible simultaneously. Previous studies did not find any performance difference between the two response modes in healthy adults (Bae, Olkkonen, Allred, Wilson, & Flombaum, 2014), and here we found no consistent evidence for reporting failures in controls, but we cannot rule out that stroke patients may be particularly susceptible to certain forms of interference which may have influenced the results.

Results from our atlas-based LSM analyses showed distinct neural profiles. We identified one lesion location associated with specific binding deficits, multiple lesion locations associated with memory precision for color, and none for location memory precision. In our behavioral data the difference between patients and controls is also more pronounced in the report-color condition compared to the report-location condition.

Critical lesions for precision in color memory in the left hemisphere are the superior parietal lobe, auditory cortex and the posterior segment of the arcuate fasciculus and the optic radiation, and in the right hemisphere the premotor/supplementary motor cortex and the inferior frontal gyrus. This pattern of a behavioral deficit as a consequence of damage to one of several brain structures is known as the equivalence brain mode and has been described in relation to memory deficits before (Godefroy et al., 1998; Toba et al., 2020). Its central principle is that behavior emerges from nodes in an organized network and that only if all nodes are intact, a 100% performance is obtained. Both the right inferior frontal gyrus and supplementary motor area have previously been associated with visuospatial working memory (Baddeley, 2003; Teramoto, Inaoka, & Ono, 2016; Xiang, Dediu, Roberts, Oort, & v., Norris, D. G., & Hagoort, P. , 2012), and with categorization (Adams & Janata, 2002; Lee et al., 2020; Li, Seger, Chen, Mo, & (March, , 2020), for colors in particular (Liu, Lu, & Seger, 2019).

In all patients with a lesion in the auditory cortex in the left hemisphere, the lesion extended to the posterior segment of the arcuate fasciculus. As our task does not have an auditory component we focus on the posterior segment of the arcuate fasciculus. The posterior segment of the arcuate fasciculus in the left hemisphere is primarily associated with the language network, specifically lexical retrieval and feedback

between visual and non-visual information (Nakajima, Kinoshita, Shinohara, & Nakada, 2019). Souza and Skóra (2017) showed that labelling of colors compared to articulatory suppression increased the quality of retention in visual working memory. A dual-content model has been proposed that distinguishes between a high-resolution channel that encodes color hues on a continuous scale, and a low-resolution channel that encodes the category of a stimulus (Bae, Olkkonen, Allred, & Flombaum, 2015). Stroke patients with a lesion including the posterior segment of the arcuate fasciculus might be impaired on verbalizing color hues and as a consequence be less accurate in reporting color. Bae and colleagues (2015) suggested that categorization in visual working memory can take place at verbal and visual level. This view is in accordance with our LSM results that associated language related areas in the left hemisphere and medial and inferior frontal areas in the right hemisphere with precision of reporting color. Our study indicates that working memory for color depends on a large anatomical network. Such a spatially distributed network is prone to damage in any brain region, which accounts for patients with different lesion locations having deficits in reporting color. However, our stroke patients demonstrated mostly subtle impairments in visual working memory. Representations from different areas in the brain might in part compensate for impaired encoding in lesioned areas.

Previous studies associated the posterior parietal cortex with the binding of features in a change detection task for shape-color bindings (Birba et al., 2017; Parra, Della Sala, Logie, & Morcom, 2014) and with perceptual attention as measured by detection of changes in color (Weber, Hahn, Hilger, & Fiebach, 2017). Results from our atlas-based LSM study identified the superior parietal lobe only for precision in the report-color condition. Binding deficits were rare in our study sample, which might explain why we did not detect an association for binding with this area.

An unexpected result was that a critical lesion location for specific binding deficits was found in the left primary somatosensory cortex (BA3). To our knowledge, to date only one study showed that visual working memory can be decoded from activity in somatosensory areas (Christophel & Haynes, 2014). This study made use of a similarity detection task for complex and colored motion stimuli. It has been suggested that representations in somatosensory areas are specific for complex dynamic stimuli (Christophel & Haynes, 2014). An alternative explanation is that the somatosensory cortex is anatomically located in the center of the fronto-parietal functional connectivity network for visual working memory (Siegel et al., 2016). Future studies should investigate the role of somatosensory areas in visual working memory.

Some of the areas we identified with a significant lesion-behavior association are typically related to motor skills or visual-motor coordination. However, for two reasons it seems unlikely that motor-related deficits, which might be a consequence of stroke, underly these associations. First, all regions were associated with only one specific outcome measure. The effect of visual-motor coordination would be expected to be comparable for the precision in reporting color and reporting location. We only found an association between precision in reporting color and a lesion in the premotor and supplementary motor cortex in the right hemisphere and the superior parietal lobe in the left hemisphere. Secondly, we found the pre-motor and supplementary motor area in the *right* hemisphere associated with precision in reporting color. Most participants were right-handed; therefore, an effect of visual-motor coordination would be expected contralaterally. The binding index is associated with the primary somatosensory cortex/postcentral gyrus in the left hemisphere, but this outcome measure is corrected for precision in reporting both color and location.

Our LSM analysis showed an association between precision in reporting color and the left optic radiation (Párraga, Ribas, Welling, Alves, & de Oliveira, 2012). While damage to the optic radiation has been associated with visual field deficits (Yogarajah et al., 2009), as far as we are aware there are no studies that associate the optic radiation with color perception or memory. All visual areas, V1-5 bilaterally, were included in our analyses but were not associated with precision in feature reporting. As most patients have unilateral lesions, the visual cortex of the intact hemisphere might compensate for possible visual deficits.

Our analysis of associations between lesion location and memory deficits is by definition limited by the lesion coverage in the tested patient population. Due to the vascularization of the brain, certain areas are unlikely to suffer a stroke (Sperber & Karnath, 2017), including some regions that have been associated with working memory. The medial temporal lobe has been implicated to play a role for selective binding deficits in Alzheimer's patients (Della Sala et al., 2012; Liang et al., 2016), with conflicting findings in other populations (see Schneegans & Bays, 2019 for a review), but this area is not typically affected by stroke (Snaphaan, Rijpkema, Uden, & v., Fernandez, G., & De Leeuw, F. E. , 2009). The dorsolateral prefrontal cortex (BA9/46) and inferior parietal lobe (BA39/40) have been reported as essential for visual working memory (Baddeley, 2003), but only areas 40 and 46 in the right hemisphere had sufficient lesion coverage to be included in the present study. A possible explanation for the small performance differences in precision in the report location condition might be the lack of lesions in key regions of the parietal lobe (e.g. Prabhakaran, Narayanan, Zhao, & Gabrieli, 2000). Based on these earlier studies and our results on visual working memory for color, we argue that memory for other features and bindings also relies on a spatially distributed network rather than on a single brain region. Representations in different regions might vary in level of abstraction, from sensory representations in primary sensory areas to abstract representations in frontal areas (Christophel, Klink, Spitzer, Roelfsema, & Haynes, 2017). Integrated representations have previously been associated with prefrontal regions, while unintegrated representations showed greater activation in posterior regions (Prabhakaran et al., 2000). A promising direction for future research is using the regions identified in the present study to identify networks underlying visual memory based on resting-state MRI in healthy subjects (Sperber & Dadashi, 2020).

To conclude, we have presented a model that explains behavioral errors in feature reporting and binding in both neurotypical controls and stroke patients. In the control group, report imprecision and swap errors in the delayed reproduction task can be explained by a single source of variability. Patients showed an overall decrease in memory precision for both color and location, but we found only limited evidence for deviations from the predicted relationship between report precision and swap errors. Binding deficits, precision in reporting color, and precision in reporting location are associated with different lesion profiles. The results from our study converge with previous reports, based on neuroimaging and other techniques, that working memory representations are widely distributed in the brain and can be found across parietal, temporal, and prefrontal cortices.

CRediT authorship contribution statement

Selma Lugtmeijer: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Data curation, Writing - original draft, Visualization. **Sebastian Schneegans:** Methodology, Software, Validation, Formal analysis, Writing - original draft, Writing - review & editing, Visualization. **Nikki A. Lammers:** Investigation, Data curation. **Linda Geerligs:** Writing - review & editing, Supervision. **Frank Erik de Leeuw:** Resources, Writing - review & editing, Supervision. **Edward H.F. de Haan:** Resources, Writing - review & editing, Supervision, Project administration, Funding acquisition. **Paul M. Bays:** Methodology, Software, Writing - review & editing, Supervision. **Roy P. C. Kessels:** Conceptualization, Writing - review & editing, Supervision.

Acknowledgements

This study was supported by an ERC Advanced Grant (#339374) awarded to E. H. F. de Haan. L. Geerligs was supported by a Veni grant

(#451-16-013) from the Netherlands Organization for Scientific Research.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.nlm.2021.107387.

References

Adams, R. B., & Janata, P. (2002). A comparison of neural circuits underlying auditory and visual object categorization. *Neuroimage, 16*, 361–377. https://doi.org/10.1006/nimg.2002.1088.

Adhikari, M. H., Hacker, C. D., Siegel, J. S., Griffa, A., Hagmann, P., Deco, G., & Corbetta, M. (2017). Decreased integration and information capacity in stroke measured by whole brain models of resting state activity. *Brain, 140*, 1068–1085. https://doi.org/10.1093/brain/awx021.

Baddeley, A. (2003). Working memory: Looking back and looking forward. *Nature Reviews Neuroscience, 4*, 829–839. https://doi.org/10.1038/nrn1201.

Bae, G. Y., Olkkonen, M., Allred, S. R., & Flombaum, J. I. (2015). Why some colors appear more memorable than others: A model combining categories and particulars in color working memory. *Journal of Experimental Psychology: General, 144*, 744–763. https://doi.org/10.1037/xge0000076.

Bae, G. Y., Olkkonen, M., Allred, S. R., Wilson, C., & Flombaum, J. I. (2014). Stimulus-specific variability in color working memory with delayed estimation. *Journal of Vision, 14*, 7. https://doi.org/10.1167/14.4.7.

Bays, P. M. (2014). Noise in neural populations accounts for errors in working memory. *Journal of Neuroscience, 34*, 3632–3645. https://doi.org/10.1523/JNEUROSCI.3204-13.2014.

Bays, P. M. (2016). Evaluating and excluding swap errors in analogue tests of working memory. *Scientific Reports, 6*, 1–14. https://doi.org/10.1038/srep19203.

Bays, P. M., Catalao, R. F., & Husain, M. (2009). The precision of visual working memory is set by allocation of a shared resource. *Journal of Vision, 9*, 7. https://doi.org/10.1167/9.10.7.

Beaudoin, A. J., Fournier, B., Julien-Caron, L., Moleski, L., Simard, J., Mercier, L., & Desrosiers, J. (2013). Visuoperceptual deficits and participation in older adults after stroke. *Australian Occupupational Therapy Journal, 60*, 260–266. https://doi.org/10.1111/1440-1630.12046.

Birba, A., Hesse, E., Sedeño, L., Mikulan, E. P., García, M.d. C., Ávalos, J., … Zimerman, M. (2017). Enhanced working memory binding by direct electrical stimulation of the parietal cortex. *Frontiers in Aging Neuroscience, 9*, 178. https://doi.org/10.3389/fnagi.2017.00178.

Bouchacourt, F., & Buschman, T. J. (2019). A flexible model of working memory. *Neuron, 103*. https://doi.org/10.1016/j.neuron.2019.04.020, 147–160. e148.

Brockmole, J. R., Parra, M. A., Della Sala, S., & Logie, R. H. (2008). Do binding deficits account for age-related decline in visual working memory? *Psychonomic Bulletin & Review, 15*, 543–547. https://doi.org/10.3758/PBR.15.3.543.

Catani, M., & De Schotten, M. T. (2008). A diffusion tensor imaging tractography atlas for virtual in vivo dissections. *Cortex, 44*, 1105–1132. https://doi.org/10.1016/j.cortex.2008.05.004.

Christophel, T. B., & Haynes, J. D. (2014). Decoding complex flow-field patterns in visual working memory. *Neuroimage, 91*, 43–51. https://doi.org/10.1016/j.neuroimage.2014.01.025.

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

Cowan, N., Naveh-Benjamin, M., Kilb, A., & Saults, J. S. (2006). Life-span development of visual working memory: When is feature binding difficult? *Developmental Psychology, 42*, 1089. https://doi.org/10.1037/0012-1649.42.6.1089.

Crinion, J., Ashburner, J., Leff, A., Brett, M., Price, C., & Friston, K. (2007). Spatial normalization of lesioned brains: Performance evaluation and impact on fMRI analyses. *Neuroimage, 37*, 866–875. https://doi.org/10.1016/j.neuroimage.2007.04.065.

De Haan, B., & Karnath, H. O. (2018). A hitchhiker's guide to lesion-behaviour mapping. *Neuropsychologia, 115*, 5–16. https://doi.org/10.1016/j.neuropsychologia.2017.10.021.

Della Sala, S., Parra, M. A., Fabi, K., Luzzi, S., & Abrahams, S. (2012). Short-term memory binding is impaired in AD but not in non-AD dementias. *Neuropsychologia, 50*, 833–840. https://doi.org/10.1016/j.neuropsychologia.2012.01.018.

DeMarco, A. T., & Turkeltaub, P. E. (2018). A multivariate lesion symptom mapping toolbox and examination of lesion-volume biases and correction methods in lesion-symptom mapping. *Human Brain Mapping, 39*, 4169–4182. https://doi.org/10.1002/hbm.24289.

Emrich, S. M., & Ferber, S. (2012). Competition increases binding errors in visual working memory. *Journal of Vision, 12*, 12. https://doi.org/10.1167/12.4.12.

Gauthier, L., Dehaut, F., & Joanette, Y. (1989). The bells test: A quantitative and qualitative test for visual neglect. *International Journal of Clinical Neuropsychology, 11*, 49–54.

Godefroy, O., Duhamel, A., Leclerc, X., Saint Michel, T., Henon, H., & Leys, D. (1998). Brain-behaviour relationships. Some models and related statistical procedures for the study of brain-damaged patients. Brain: *A. Journal of Neurology, 121*, 1545–1556. https://doi.org/10.1093/brain/121.8.1545.

Lee, J. K., Fandakova, Y., Johnson, E. G., Cohen, N. J., Bunge, S. A., & Ghetti, S. (2020). Changes in anterior and posterior hippocampus differentially predict item-space, item-time, and item-item memory improvement. *Developmental Cognitive Neuroscience, 41*, 100741. https://doi.org/10.1016/j.dcn.2019.100741.

Li, Y., Seger, C., Chen, Q., & Mo, L. (2020). Left Inferior frontal Gyrus integrates multisensory information in category learning. *Cerebral Cortex*. https://doi.org/10.1093/cercor/bhaa029.

Liang, Y., Pertzov, Y., Nicholas, J. M., Henley, S. M., Crutch, S., Woodward, F., … Husain, M. (2016). Visual short-term memory binding deficit in familial Alzheimer's disease. *Cortex, 78*, 150–164. https://doi.org/10.1016/j.cortex.2016.01.015.

Liu, Z., Lu, W., & Seger, C. A. (2019). Perceptual and categorical processing and representation in color categorization. *Brain and Cognition, 136*, 103617. https://doi.org/10.1016/j.bandc.2019.103617.

Matthey, L., Bays, P. M., & Dayan, P. (2015). A probabilistic palimpsest model of visual short-term memory. *PLoS Computational Biology, 11*. https://doi.org/10.1371/journal.pcbi.1004003.

Mayes, A., Montaldi, D., & Migo, E. (2007). Associative memory and the medial temporal lobes. *Trends in Cognitive Sciences, 11*, 126–135. https://doi.org/10.1016/j.tics.2006.12.003.

Mitchell, K. J., Johnson, M. K., Raye, C. L., Mather, M., & D'Esposito, M. (2000). Aging and reflective processes of working memory: Binding and test load deficits. *Psychology and Aging, 15*, 527. https://doi.org/10.1037/0882-7974.15.3.527.

Moses, S. N., & Ryan, J. D. (2006). A comparison and evaluation of the predictions of relational and conjunctive accounts of hippocampal function. *Hippocampus, 16*, 43–65. https://doi.org/10.1002/hipo.20131.

Nakajima, R., Kinoshita, M., Shinohara, H., & Nakada, M. (2019). The superior longitudinal fascicle: Reconsidering the fronto-parietal neural network based on anatomy and function. *Brain Imaging and Behavior, 1–14*. https://doi.org/10.1007/s11682-019-00187-4.

Oberauer, K., & Lin, H. Y. (2017). An interference model of visual working memory. *Psychological Review, 124*, 21. https://doi.org/10.1037/rev0000044.

Olson, I. R., Page, K., Moore, K. S., Chatterjee, A., & Verfaellie, M. (2006). Working memory for conjunctions relies on the medial temporal lobe. *Journal of Neuroscience, 26*, 4596–4601. https://doi.org/10.1523/JNEUROSCI.1923-05.2006.

Parra, M. A., Della Sala, S., Logie, R. H., & Morcom, A. M. (2014). Neural correlates of shape–color binding in visual working memory. *Neuropsychologia, 52*, 27–36. https://doi.org/10.1016/j.neuropsychologia.2013.09.036.

Parra, M. A., Fabi, K., Luzzi, S., Cubelli, R., Hernandez Valdez, M., & Della Sala, S. (2015). Relational and conjunctive binding functions dissociate in short-term memory. *Neurocase, 21*, 56–66. https://doi.org/10.1080/13554794.2013.860177.

Parra, M. A., Saarimäki, H., Bastin, M. E., Londoño, A. C., Pettit, L., Lopera, F., … Abrahams, S. (2015). Memory binding and white matter integrity in familial Alzheimer's disease. *Brain, 138*, 1355–1369. https://doi.org/10.1093/brain/awv048.

Párraga, R. G., Ribas, G. C., Welling, L. C., Alves, R. V., & de Oliveira, E. (2012). Microsurgical anatomy of the optic radiation and related fibers in 3-dimensional images. Operative. *Neurosurgery, 71*(suppl_1). ons160-ons172.

Pertzov, Y., Heider, M., Liang, Y., & Husain, M. (2015). Effects of healthy ageing on precision and binding of object location in visual short term memory. *Psychology and Aging, 30*, 26. https://doi.org/10.1037/a0038396.

Pertzov, Y., Miller, T. D., Gorgoraptis, N., Caine, D., Schott, J. M., Butler, C., & Husain, M. (2013). Binding deficits in memory following medial temporal lobe damage in patients with voltage-gated potassium channel complex antibody-associated limbic encephalitis. *Brain, 136*, 2474–2485. https://doi.org/10.1093/brain/awt129.

Piekema, C., Rijpkema, M., Fernández, G., & Kessels, R. P. (2010). Dissociating the neural correlates of intra-item and inter-item working-memory binding. *PLoS One, 5*, e10214. https://doi.org/10.1371/journal.pone.0010214.

Prabhakaran, V., Narayanan, K., Zhao, Z., & Gabrieli, J. D. E. (2000). Integration of diverse information in working memory within the frontal lobe. *Nature Neuroscience, 3*, 85–90. https://doi.org/10.1038/71156.

Rajsic, J., & Wilson, D. E. (2014). Asymmetrical access to color and location in visual working memory. *Attention, Perception, & Psychophysics, 76*, 1902–1913. https://doi.org/10.3758/s13414-014-0723-2.

Rerko, L., Oberauer, K., & Lin, H. Y. (2014). Spatial transposition gradients in visual working memory. *The Quarterly Journal of Experimental Psychology, 67*, 3–15. https://doi.org/10.1080/17470218.2013.789543.

Rhodes, S., Parra, M. A., & Logie, R. H. (2016). Ageing and feature binding in visual working memory: The role of presentation time. *The Quarterly Journal of Experimental Psychology, 69*, 654–668. https://doi.org/10.1080/17470218.2015.1038571.

Rorden, C., Bonilha, L., Fridriksson, J., Bender, B., & Karnath, H. O. (2012). Age-specific CT and MRI templates for spatial normalization. *Neuroimage, 61*, 957–965. https://doi.org/10.1016/j.neuroimage.2012.03.020.

Rowe, F. J., Group, V. W., Wright, D., Brand, D., Maan, T., Peel, S., … Shipman, T. (2017). Vision In Stroke cohort: Profile overview of visual impairment. *Brain and Behavior, 7*. https://doi.org/10.1002/brb3.771.

Schmand, B., Ooms, M., & Ribbe, M. (1997). Abbreviated form of the Informant Questionnaire on cognitive decline in the elderly. *Tijdschrift voor gerontologie en geriatrie, 28*, 224–229.

Schneegans, S., & Bays, P. M. (2017). Neural architecture for feature binding in visual working memory. *Journal of Neuroscience, 37*, 3913–3925. https://doi.org/10.1523/JNEUROSCI.3493-16.2017.

Schneegans, S., & Bays, P. M. (2019). New perspectives on binding in visual working memory. *British Journal of Psychology, 110*, 207–244. https://doi.org/10.1111/bjop.12345.

Schneegans, S., Taylor, R., & Bays, P. (2020). Stochastic sampling provides a unifying account of working memory limits. *BioRxiv, 771071*. https://doi.org/10.1101/771071.

Siegel, J. S., Ramsey, L. E., Snyder, A. Z., Metcalf, N. V., Chacko, R. V., Weinberger, K., … Corbetta, M. (2016). Disruptions of network connectivity predict impairment in multiple behavioral domains after stroke. *Proceedings of the National Academy of Sciences, 113*, E4367–E4376. https://doi.org/10.1073/pnas.1521083113.

Snaphaan, L., Rijpkema, M., Uden, I. V., Fernandez, G., & De Leeuw, F. E. (2009). Reduced medial temporal lobe functionality in stroke patients: A functional magnetic resonance imaging study. *Brain, 132*, 1882–1888. https://doi.org/10.1093/brain/awp133.

Souza, A. S., & Skóra, Z. (2017). The interplay of language and visual perception in working memory. *Cognition, 166*, 277–297. https://doi.org/10.1016/j.cognition.2017.05.038.

Sperber, C., & Dadashi, A. (2020). The influence of sample size and arbitrary statistical thresholds in lesion-network mapping. *Brain, 5*, Article e40. https://doi.org/10.1093/brain/awaa094.

Sperber, C., & Karnath, H. O. (2017). Impact of correction factors in human brain lesion-behavior inference. *Human Brain Mapping, 38*, 1692–1701. https://doi.org/10.1002/hbm.23490.

Teramoto, S., Inaoka, T., & Ono, Y. (2016). Regional brain activity that determines successful and unsuccessful working memory formation. Paper presented at the 2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). Orlando. *FL, 2016*, 5550–5553. https://doi.org/10.1109/EMBC.2016.7591984.

Toba, M. N., Godefroy, O., Rushmore, R. J., Zavaglia, M., Maatoug, R., Hilgetag, C. C., & Valero-Cabré, A. (2020). Revisiting 'brain modes' in a new computational era: Approaches for the characterization of brain-behavioural associations. *Brain, 4*, 1088–1098. https://doi.org/10.1093/brain/awz343.

Weber, E. M. G., Hahn, T., Hilger, K., & Fiebach, C. J. (2017). Distributed patterns of occipito-parietal functional connectivity predict the precision of visual working memory. *Neuroimage, 146*, 404–418. https://doi.org/10.1016/j.neuroimage.2016.10.006.

Wilken, P., & Ma, W. J. (2004). A detection theory account of change detection. *Journal of Vision, 4*, 11. https://doi.org/10.1167/4.12.11.

Xiang, H., Dediu, D., Roberts, L., Oort, E. V., Norris, D. G., & Hagoort, P. (2012). The structural connectivity underpinning language aptitude, working memory, and IQ in the perisylvian language network. *Language Learning, 62*, 110–130. https://doi.org/10.1111/j.1467-9922.2012.00708.x.

Yogarajah, M., Focke, N., Bonelli, S., Cercignani, M., Acheson, J., Parker, G., … Koepp, M. (2009). Defining Meyer's loop–temporal lobe resections, visual field deficits and diffusion tensor tractography. *Brain, 132*, 1656–1668. https://doi.org/10.1093/brain/awp114.

Yushkevich, P. A., Piven, J., Hazlett, H. C., Smith, R. G., Ho, S., Gee, J. C., & Gerig, G. (2006). User-guided 3D active contour segmentation of anatomical structures: Significantly improved efficiency and reliability. *Neuroimage, 31*, 1116–1128. https://doi.org/10.1016/j.neuroimage.2006.01.015.

Consequence of stroke for feature recall and binding in visual working memory
Supplementary material

Selma Lugtmeijer[1,2], Sebastian Schneegans[3], Nikki A. Lammers[1], Linda Geerligs[2], Frank Erik de Leeuw[4], Edward H. F. de Haan[1], Paul M. Bays[3], and Roy P. C. Kessels[2,5]

[1]University of Amsterdam, The Netherlands
[2]Donders Institute for Brain, Cognition and Behaviour, Radboud University, Nijmegen, The Netherlands
[3]University of Cambridge, Department of Psychology, Cambridge, United Kingdom
[4]Radboud University Medical Center, Department of Neurology, Nijmegen, The Netherlands
[5]Radboud University Medical Center, Department of Medical Psychology, Nijmegen, The Netherlands

1 Data analysis

We denote the report feature values of the N sample items in trial i as $\{\theta_1^{(i)}, \ldots, \theta_N^{(i)}\}$, with $\theta_1^{(i)}$ being the target feature value, and we denote the response feature value as $\theta_r^{(i)}$. The response errors are then determined as

$$\epsilon^{(i)} = D_\circ(\theta_r^{(i)}, \theta_1^{(i)}), \tag{1}$$

and the non-target deviations as

$$\tilde{\epsilon}_j^{(i)} = D_\circ(\theta_r^{(i)}, \theta_j^{(i)}) \quad \text{for} \quad j = 2, \ldots, N, \tag{2}$$

where $D_\circ$ is the signed distance on the circle.

The occurrence of swap errors can be visualized by plotting the histogram of non-target deviations, with a central peak indicating that responses are

clustered around the report feature values of non-target items. However, if there is a minimum distance between the feature values of all sample items within a trial (as is the case in the present experiment), the distribution of non-target deviations cannot be assumed to be uniform in the absence of swap errors. If the response values are concentrated around the target value, they will tend to be at least that minimum distance away from the report values of the non-target items, resulting in a central dip in the distribution of non-target deviations which may mask any central peak produced by swap errors.

We therefore correct the histogram of non-target deviations by subtracting the expected histogram in the absence of any swap errors (Schneegans and Bays, 2017), computed separately for each participant and each task condition. We determine the deviation of all non-target features from the target feature in each trial,

$$\delta_j^{(i)} = D_\circ(\theta_j^{(i)}, \theta_1^{(i)}) \quad \text{for} \quad j = 2, \ldots, N, \tag{3}$$

and then compute the histogram over all differences

$$\zeta_j^{(i,i')} = D_\circ(\epsilon^{(i)}, \delta_j^{(i')}) \quad \text{for} \quad j = 2, \ldots, N \quad \text{and} \quad i, i' = 1, \ldots, T, \tag{4}$$

where T is the number of trials in each condition. This yields the expected histogram of non-target deviations by shuffling the deviations of responses from targets and the relative position of non-targets to targets across trials.

To test for the presence of swap errors, we determined for each participant the arithmetic mean of the absolute non-target deviations, $|\tilde{\epsilon}_j^{(i)}|$, across all non-targets and trials, and the mean of all shuffled absolute non-target deviations, $|\zeta_j^{(i,i')}|$, and compared these using a paired-samples t-test.

2 Neural binding model

2.1 Conjunctive population code

We assume that the colors and locations of the sample stimuli are encoded in an idealized conjunctive population code, in which each neuron's activity is determined by its tuning functions for stimulus color and location. Recall is modeled as decoding of memorized features from noisy neural activity. We will describe this neural population model in terms of cue and report feature values. Either role can be taken by color or location, depending on task condition.

The firing rate of neuron k encoding cue feature ψ_j and report feature θ_j of item j in the sample display is given as

$$\bar{r}_{k,j}(\psi_j, \theta_j) = \frac{\gamma}{NM}\phi_\circ\big(\psi_j; \psi'_k, \kappa_\psi\big)\phi_\circ\big(\theta_j; \theta'_k, \kappa_\theta\big) \tag{5}$$

Here, γ is the mean total firing rate of the population, which is divided by the number of sample items, N, and the number of neurons, M, that contribute to the encoding of each item. The feature tuning of the neuron is described by von Mises functions with preferred values ψ'_k and θ'_k for cue and report feature, respectively, and associated concentration parameters κ_ψ and κ_θ. We assume that the shape of the tuning curves is fixed throughout the population, and individual neurons only differ in their preferred feature values, which uniformly sample the underlying feature space of color-location combinations.

Discrete spikes are produced based on each neuron's firing rate via independent Poisson processes,

$$r_{k,j} \sim \text{Pois}(\bar{r}_{k,j}) \tag{6}$$

Due to the superposition property of the Poisson distribution, the total number of spikes, n_j, that contribute to representing the features of each item j is then likewise a Poisson random variable,

$$n_j \sim \text{Pois}\Big(\frac{\gamma}{N}\Big). \tag{7}$$

2.2 Response probabilities

Feature recall is modeled as maximum likelihood estimation of the encoded feature values from the noisy spiking activity over a fixed time window. To determine the distribution of decoding errors, we deviate from the method used by Schneegans and Bays (2017), and instead build on new results from Schneegans et al. (2019) to derive a more elegant solution. In this publication it has been shown that for a given number of spikes contributing to the encoding of item j, the distribution of decoded values $\hat{\theta}_j$ can be closely approximated by a von Mises distribution around the true feature value θ_j in each feature dimension, with precision scaled by the number of spikes n_j:

$$p_{\text{dec}}\left(\hat{\theta}_j \;\middle|\; \theta_j, n_j\right) = \phi_\circ\Big(\hat{\theta}_j; \theta_j, \kappa(n_j\omega_\theta)\Big) \tag{8}$$

Here, ω_θ is the precision (as Fisher information) corresponding to the tuning curve concentration κ_θ, which is determined as $\omega = \kappa\frac{I_1(\kappa)}{I_0(\kappa)}$, and the term

$\kappa(n_j\omega_\theta)$ describes the concentration parameter yielding a multiple of the base precision ω_θ, which can be obtained by numerical inversion of the same relationship.

The joint distribution of decoded cue and report feature values can then be described as

$$p_{\text{dec}}\left(\hat{\theta}_j, \hat{\psi}_j \;\middle|\; \theta_j, \psi_j\right) = \sum_{n_j=0}^{\infty} \Pr_{\text{Pois}}\left(n_j, \frac{\gamma}{N}\right) p_{\text{dec}}\left(\hat{\theta}_j \;\middle|\; \theta_j, n_j\right) p_{\text{dec}}\left(\hat{\psi}_j \;\middle|\; \psi_j, n_j\right) \quad (9)$$

It should be noted that decoding errors in the two feature dimensions are not independent of each other, since both depend on the number of spikes in the neural population that contribute to decoding the item's features.

We assume that the cue and response features of all items are decoded from the neural activity when a cue is given. The item whose decoded cue feature value is closest to the actual cue is selected for response generation, and its decoded report feature value is produced as response. The probability that a certain report feature value θ_r is chosen as a response in a trial with item report and cue feature values $\boldsymbol{\theta}$ and $\boldsymbol{\psi}$, respectively, is then

$$p_{\text{resp}}\left(\theta_r \mid \boldsymbol{\theta}, \boldsymbol{\psi}\right) = \sum_{j=1}^{N} p\left(\hat{\theta}_j = \theta_r \wedge \text{item } j \text{ selected} \;\middle|\; \boldsymbol{\theta}, \boldsymbol{\psi}\right). \quad (10)$$

The probability that an item is selected for response generation is determined by its decoded cue feature, and due to the aforementioned dependence between decoding errors it is not independent from the obtained report feature value. But we can separate these probabilities by conditioning on the number of available spikes, n_j:

$$p_{\text{resp}}\left(\theta_r \mid \boldsymbol{\theta}, \boldsymbol{\psi}\right) = \sum_{j=1}^{N} \sum_{n_j=0}^{\infty} \Pr_{\text{Pois}}\left(n_j, \frac{\gamma}{N}\right) p_{\text{dec}}\left(\theta_r \mid \theta_j, n_j\right) \Pr_{\text{sel}}\left(j \mid \boldsymbol{\psi}, n_j\right) \quad (11)$$

The conditional probability of decoding a certain report feature value given the spike count and true feature value in this equation can be determined as in Eq. 8.

The probability that an item is selected (i.e., its decoded cue feature value is closest to the actual cue) can be computed by numerical integration

as

$$\begin{aligned}&\Pr_{\text{sel}}\left(j \mid \boldsymbol{\psi}, n_j\right)\\&\quad= \int_0^{\pi} p\left(D_\circ(\hat{\psi}_j - \psi_c) = s \;\middle|\; \psi_j, n_j\right) \prod_{j' \neq j} p\left(D_\circ(\hat{\psi}_{j'} - \psi_c) > s \;\middle|\; \psi_{j'}\right) ds,\end{aligned} \tag{12}$$

where ψ_c is the feature value of the actually given cue. The first probability term in this integral can be evaluated based on Eq. 8, while the second term requires marginalizing over the possible sample counts,

$$p_{\text{dec}}\left(\hat{\psi}_{j'} \;\middle|\; \psi_{j'}\right) = \sum_{n_{j'}=0}^{\infty} \Pr_{\text{Pois}}\left(n_{j'}, \frac{\gamma}{N}\right) p_{\text{dec}}\left(\hat{\psi}_{j'} \;\middle|\; \psi_{j'}, n_{j'}\right). \tag{13}$$

2.3 Binding and reporting deficits

In order to detect specific impairments in feature binding performance, we extend the model in a way which relaxes the assumption that memory precision for a feature when used as a cue from is the same as memory precision for the same feature when it is reported. More specifically, we allow the number of spikes that contribute to the selection of the cued item to be different from the number of spikes that contribute to decoding of the report features. This is compatible with the idea that the pool of neurons underlying memory for individual features may be separate from the one underlying binding memory, without making any strong assumptions about the specific neural architecture.

We introduce a new parameter a_{select} that specifies the mean proportion of total spikes n_j that are available for selecting an item for response based on the cue. We assume that this adjusted number of spikes $\tilde{n}_j$ is drawn from a binomial distribution with success rate a_{select}, such that the selection probability used in Eq. 11 is now given as

$$\Pr_{\text{sel}}\left(j \mid \boldsymbol{\psi}, n_j, a_{\text{select}}\right) = \sum_{\tilde{n}_j=0}^{n_j} \Pr_{\text{Binom}}(\tilde{n}_j; n_j, a_{\text{select}}) \Pr_{\text{sel}}\left(j \mid \boldsymbol{\psi}, \tilde{n}_j\right), \tag{14}$$

where $\Pr_{\text{sel}}\left(j \mid \boldsymbol{\psi}, \tilde{n}_j\right)$ is again determined as in Eq. 12.

We also allow for the converse effect, i.e. an impairment of reporting the feature value after an item has been selected. For this case, we assume that the number of spikes for decoding the report feature is a subset of the total

spikes, likewise drawn from a binomial distribution with success rate a_{report}. The decoding probability of the report in Eq. 11 is then computed as

$$p_{\text{dec}}\left(\theta \mid \theta_j, n_j, a_{\text{report}}\right) = \sum_{\tilde{n}_j=0}^{n_j} \Pr_{\text{Binom}}(\tilde{n}_j; n_j, a_{\text{report}}) p_{\text{dec}}\left(\theta \mid \theta_j, \tilde{n}_j\right), \quad (15)$$

with $p_{\text{dec}}\left(\theta \mid \theta_j, \tilde{n}_j\right)$ determined as in Eq. 8.

We combine the model variants with binding deficit and reporting deficit into a single model with a *binding index* b as free parameter, in such a way that $b = 0$ reflects no binding or reporting deficit (all spikes are available both for selecting the report item and decoding its report feature value), $b = -1$ indicates complete loss of binding memory (no spikes available for selecting the report item, so each sample item is selected with equal probability) and $b = 1$ indicates complete loss of feature reporting ability (no spikes available for decoding the report feature value, so all responses are drawn from a uniform distribution):

$$\begin{array}{ll} a_{\text{select}} = 1 + b,\ a_{\text{report}} = 1 & \text{if } b \leq 0 \\ a_{\text{select}} = 1,\ a_{\text{report}} = 1 - b & \text{otherwise} \end{array} \quad (16)$$

2.4 Priors for model parameters

Due to the very limited amount of behavioral data collected for each participant, some aspects of the model fits can be underconstrained in the current study. The first of these concerns a trade-off between the mean total spike rate γ and the tuning curve concentrations κ_θ and κ_ψ. An increase in recall precision can be achieved in the model either by increasing the spike rate or the concentration parameters. In most VWM studies, recall performance is measured at different set sizes. The neural population model assumes that the total spike rate is distributed among all sample items, while the tuning curves remain fixed across set sizes. This mechanism has been shown to successfully account for set size effects (Bays, 2014), and provides sufficient constraints to obtain robust estimates for each parameter.

In the present study with a single set size and small number of trials, we employ a weakly informative prior on the parameter γ. The prior is based on population model fits to a database of delayed reproduction tasks with color report (Schneegans et al., 2019), but with increased variability to avoid overly constraining the model fits. It is implemented as a Gamma distribution,

$$p(\gamma) = \frac{1}{\Gamma(k)\theta^k} \gamma^{k-1} e^{\frac{\gamma}{\theta}}, \quad (17)$$

with shape parameter $k = 2$ and scale parameter $\theta = 8$. This prior penalizes extremely small values of γ as well as very large values. In particular it prevents γ from going towards infinity in the model fits (while the κ values go towards zero), which otherwise happens for a few participants, without substantially altering the resulting error distributions.

Another issue that arises in fitting the model to the data is that some participants do not show any identifiable swap errors, due to the small number of trials and the relatively low difficulty of the task. In these participants, increasing the precision for the cue feature towards infinity improves the quality of fit in each condition. To avoid unrealistically high estimates of cue feature precision, we add a weakly informative prior on the probability of swap errors. This prior is implemented by computing for a given set of model parameters the probability that a swap error occurs if both non-target items have the minimum allowed distance (30°) to the target in the cue dimension, using Eq. 12. Then a Beta-distribution distribution is applied to this probablity p_{NT},

$$p(p_{\mathrm{NT}}) = \frac{p_{\mathrm{NT}}^{\alpha-1}(1-p_{\mathrm{NT}})^{\beta-1}}{B(\alpha,\beta)} \tag{18}$$

with $\alpha = \beta = 2$. This prior is directly equivalent to adding two trials with minimum distance between cue features to each participant's data in each condition, one of which results in a swap error and the other in a target response (while ignoring the actually reported feature), and it penalizes both very small and very high (close to one) swap probabilities.

2.5 Model fitting and comparison

We determined maximum likelihood fits of each model to the behavioral data of each participant. For the neural binding model, we obtained both separate fits for each task condition (six parameters in total), and a combined fit with shared parameters across both condition (parameters γ, κ_{color} and $\kappa_{\mathrm{location}}$, with the latter assigned either to the cue or the report dimension according to task condition). The model with additional binding index, b, was fit to the combined data only (four parameters in total). Maximum likelihood fits were determined via the Nelder-Mead simplex method (function `fminsearch` in Matlab), using a grid of possible initial values for all parameters. Initial values were $[6, 12, 24]$ for γ, $[2, 5, 12]$ for κ in each feature dimension, and $[-0.3, 0, 0.3]$ for b.

We compare models' quality of fit using the Akaike information criterion

with correction for small sample size (AICc),

$$AICc = 2k - 2\ln L + \frac{2k^2 + 2k}{n - k - 1}, \tag{19}$$

where k is the total number of free parameters in each model, L is the likelihood of the fitted model, and n is the total number of trials for each participant.

The pattern of results would not qualitatively change if we used the Bayesian information criterion instead of the AICc for model comparisons, although the combined fit of both task conditions with the original neural binding model (which has the lowest number of free parameters) would have an even larger advantage over the alternative models.

2.6 Model-based performance measures

We use the circular standard deviation of the decoding errors in the absence of binding or reporting deficits as a measure of memory performance. To this end, we compute the probability distribution $p_{dec}(\hat{\theta} \,|\, \theta)$ as in Eq. 13, and numerically determine its circular standard deviation. This measure incorporates the concentration parameters of the tuning curves in each feature dimension, κ_{location} and κ_{color}, as well as the shared spike rate parameter γ. Due to the possible trade-off between these parameters described above, we do not analyze and compare these individual parameters directly. Additionally, we use the binding index estimated for each participant as measure of specific binding or reporting deficits.

We can also estimate the proportion of swap errors that occur for each participant from the model fits. For a single trial, the posterior probability that the given response θ_r was the result of selecting item j for response generation can be derived from Eq. 11 as

$$\Pr\left(j \mid \theta_r, \boldsymbol{\theta}, \boldsymbol{\psi}\right) = \frac{\sum_{n_j=0}^{\infty} \Pr_{\text{Pois}}\left(n_j, \frac{\gamma}{N}\right) p_{\text{dec}}\left(\theta_r \mid \theta_j, n_j\right) \Pr_{\text{sel}}\left(j \mid \boldsymbol{\psi}, n_j\right)}{p_{\text{resp}}\left(\theta_r \mid \boldsymbol{\theta}, \boldsymbol{\psi}\right)}. \tag{20}$$

To estimate the overall proportion of swap errors, we sum the probability values obtained from this equation for the two non-target items in each trial, and average the sum over all trials.

3 Lesion coverage

Table S1. Number of participants per region of interest. Shaded areas were included in the atlas-based LSM analysis. Only the areas in a dark shade of grey were significantly associated with one of the outcome measures. The number of subjects indicates how many subjects had a lesion in each area.

Brodmann areas		***N*** subjects
1_L	primary somatosensory cortex	3
1_R	primary somatosensory cortex	2
2_L	primary somatosensory cortex	3
2_R	primary somatosensory cortex	4
3_L	primary somatosensory cortex	7
3_R	primary somatosensory cortex	5
4_L	primary motor cortex	8
4_R	primary motor cortex	5
5_L	somatosensory association cortex	2
5_R	somatosensory association cortex	3
6_L	premotor cortex and supplementary motor cortex	15
6_R	premotor cortex and supplementary motor cortex	11
7_L	superior parietal lobe	4
7_R	superior parietal lobe	5
8_L	frontal eyefield	2
8_R	frontal eyefield	3
9_L	dorsolateral prefrontal cortex	2
9_R	dorsolateral prefrontal cortex	3
10_L	anterior prefrontal cortex	1
10_R	anterior prefrontal cortex	1
11_L	orbitofrontal	2
11_R	orbitofrontal	3
17_L	V1	6
17_R	V1	8
18_L	V2	7
18_R	V2	10
19_L	V3,4,5	10
19_R	V3,4,5	12
20_L	inferior temporal gyrus	3
20_R	inferior temporal gyrus	4
21_L	middle temporal gyrus	4

21_R	middle temporal gyrus	4
22_L	superior temporal gyrus	3
22_R	superior temporal gyrus	5
23_L	cingulate cortex	2
23_R	cingulate cortex	2
24_L	cingulate cortex	1
24_R	cingulate cortex	1
25_L	subgenual area	0
25_R	subgenual area	2
26_L	retrosplenial region	0
26_R	retrosplenial region	0
27_L	piriform cortex	0
27_R	piriform cortex	2
28_L	cingulate cortex	0
28_R	cingulate cortex	1
29_L	cingulate cortex	0
29_R	cingulate cortex	0
30_L	cingulate cortex	1
30_R	cingulate cortex	4
32_L	cingulate cortex	0
32_R	cingulate cortex	2
34_L	dorsal enthorihinal cortex	2
34_R	dorsal enthorihinal cortex	2
35_L	perirhinal cortex	1
35_R	perirhinal cortex	2
36_L	perirhinal cortex	0
36_R	perirhinal cortex	1
37_L	fusiform gyrus	6
37_R	fusiform gyrus	9
38_L	temporal pole	3
38_R	temporal pole	4
39_L	angular gyrus	3
39_R	angular gyrus	3
40_L	supramarginal gyrus	3
40_R	supramarginal gyrus	5
41_L	primary auditory cortex / heschl gyrus	5
41_R	primary auditory cortex / heschl gyrus	3
42_L	primary auditory cortex / heschl gyrus	2
42_R	primary auditory cortex / heschl gyrus	4
43_L	primary gustatory cortex	6

43_R	primary gustatory cortex	4
44_L	broca	10
44_R	broca	9
45_L	broca	4
45_R	broca	5
46_L	dorsolateral prefrontal cortex	3
46_R	dorsolateral prefrontal cortex	5
47_L	inferior frontal gyrus	4
47_R	inferior frontal gyrus	6
48_L	retrosubicular area	25
48_R	retrosubicular area	18

CAT	atlas	***N*** subjects
1	anterior commissure left	2
2	arcuate anterior segment left	11
3	long segment left	13
4	arcuate posterior segment left	6
5	cingulum left	9
6	corpus callosum left	19
7	cortico-ponto cerebellum left	20
8	cortico-spinal left	27
9	fornix left	3
10	inferior cerebellar pedunculus left	4
11	inferior longitudinal fasciculus left	11
12	inferior occipito-frontal fasciculus left	11
13	internal capsule left	25
14	optic radiations left	13
15	superior cerebelar pedunculus left	2
16	uncinate left	7
17	anterior commissure right	5
18	arcuate anterior segment right	9
19	long segment right	6
20	arcuate posterior segment right	5
21	cingulum right	9
22	corpus callosum right	21
23	cortico-ponto cerebellum right	13
24	cortico-spinal right	18
25	fornix right	5
26	inferior cerebellar pedunculus right	2
27	inferior longitudinal fasciculus right	11

28	inferior occipito-frontal fasciculus right	18
29	internal capsule right	21
30	optic radiations right	7
31	superior cerebelar pedunculus right	2
32	uncinate right	9

References

Bays, P. M. (2014). Noise in neural populations accounts for errors in working memory. *Journal of Neuroscience*, 34(10):3632–3645.

Schneegans, S. and Bays, P. M. (2017). Neural architecture for feature binding in visual working memory. *Journal of Neuroscience*, 37(14):3913–3925.

Schneegans, S., Taylor, R., and Bays, P. M. (2019). Stochastic sampling provides a unifying account of working memory limits. *BioRxiv*, page 771071.