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.