Published in final edited form as:
J Cogn Neurosci. 2022 March 31; 34(5): 776–786. doi:10.1162/jocn_a_01831.

Working memory swap errors have identifiable neural representations

Remington Mallett,
Elizabeth S. Lorenc,
Jarrod A. Lewis-Peacock[*]
Department of Psychology, University of Texas at Austin

Abstract

Working memory is an essential component of cognition that facilitates goal-directed behavior. Famously, it is severely limited and performance suffers when memory load exceeds an individual's capacity. Modeling of visual working memory responses has identified two likely types of errors: guesses and swaps. Swap errors may arise from a mis-binding between the features of different items. Alternatively, these errors could arise from memory noise in the feature dimension used for cueing a to-be-tested memory item, resulting in the wrong item being selected. Finally, it is possible that so-called "swap errors" actually reflect informed guessing, which could occur at the time of a cue, or alternatively, at the time of the response. Here, we combined behavioral response modeling and fMRI pattern analysis to test the hypothesis that swap errors involve the active maintenance of an incorrect memory item. After the encoding of six spatial locations, a retro-cue indicated which location would be tested after memory retention. On accurate trials, we could reconstruct a memory representation of the cued location in both early visual cortex and intraparietal sulcus. On swap error trials identified with mixture modeling, we were able to reconstruct a representation of the swapped location, but not of the cued location, suggesting the maintenance of the incorrect memory item prior to response. Moreover, participants subjectively responded with some level of confidence, rather than complete guessing, on a majority of swap error trials. Together, these results suggest that swap errors are not mere response-phase guesses, but instead result from failures of selection in working memory, contextual binding errors, or informed guesses, which produce active maintenance of incorrect memory representations.

Keywords

short-term memory; binding errors; mixture model; fMRI

Introduction

Working memory serves to aid in the execution of short-term behavioral goals (Miller et al., 1960), be that for action planning (Myers et al., 2017) or the stabilization of perception into a coherent experience (Kiyonaga et al., 2017). Importantly, there are limitations within

[*] jalewpea@utexas.edu .

and across individuals as to how much information can be maintained and manipulated in working memory (Luck & Vogel, 2013; Oberauer et al., 2016). When working memory capacity is exceeded, errors occur. Traditionally, this capacity limitation has been quantified as percent-correct across multiple trials (e.g., Luck & Vogel, 1997), though the application of continuous response measures to working memory tasks (e.g., Wilken & Ma, 2004; W. Zhang & Luck, 2008) radically shifted the field towards behavioral response models that quantify more subtle and differentiable response profiles (Ma et al., 2014). Such modeling of working memory behavior has shown that responses can be largely described by a mixture distribution composed of one uniform component (representing guess responses) and two normal distribution components (representing accurate and swap error responses) (Bays et al., 2009). This analysis method, referred to as mixture modeling, has since been widely implemented, extended, and challenged (Schurgin et al., 2020).

Guess responses involve an incorrect response that matches neither the target nor any of the non-target memory items. They are believed to simply result from the absence of the relevant memory information. In contrast, swap errors involve an incorrect response *that matches one of the non-target memory items*. The interpretation of swap errors is less straightforward. On some trials, they may simply be guesses that happen to match one of the other memory items. However, such non-target responses are frequently observed at a higher rate than would be predicted by random guessing, suggesting that they are a distinct phenomenon that arises from feature binding failures in working memory (Bays et al., 2009). Working memory items typically consist of multiple features (e.g., a color and a location), one of which will be tested and the other will be used for selection. Swap errors may result from a misbinding binding between these features (Treisman, 1996; Treisman & Schmidt, 1982; Wheeler & Treisman, 2002) occurring either during initial encoding (Golomb, 2015; Golomb et al., 2014) or across the memory delay (Bays, 2015; Pertzov et al., 2012; Schneegans & Bays, 2017). Alternatively, noisy memory representations in the feature dimension used for selection could cause an incorrect item (and its bound location feature value) to be selected (Bays, 2009; Bays, 2015; Schneegans & Bays, 2017; Schneegans & Bays, 2019). In either case, this leads participants to report the location of an incorrect item from the memory set. The rate at which swap errors occur varies widely, including estimates of 16% (van den Berg et al., 2014) and 30% (Bays, 2016) of all trials, and such confusion errors are considered a “benchmark” trait of working memory (Oberauer et al., 2018).

Most theoretical discussion of swap errors is built upon behavioral data and hypothetical neural modeling (Schneegans & Bays, 2019). The lack of neuroimaging evidence for the existence of swap errors has motivated some to challenge the construct altogether in favor of a view in which all errors derive from guessing (Huang, 2020; Pratte, 2018). In the current experiment, we set out to test whether swap errors involve the selection and retention of a valid, but incorrect memory item during a working memory delay period. We designed a behavioral task to maximize swap error rates and measured participants’ brain activity with functional magnetic resonance imaging (fMRI) while they completed ≥200 trials across two scanning sessions. To maximize detection of mnemonic neural activity, the task involved maintenance of specific spatial locations of colored circles, which is known to recruit early visual cortex (Awh et al., 2000; Blacker & Courtney, 2016; Brignani et al., 2010; Munneke et al., 2010; Peters et al., 2015; Pratte & Tong, 2014; Sprague et al., 2014). We used a multivariate inverted encoding model (IEM; Sprague et al., 2014, 2016) to reconstruct spatial memory representations throughout the working memory delay period on each trial. Swap errors were identified using trial-level behavioral modeling (Schneegans & Bays, 2016), and we found that both early visual cortex and intraparietal sulcus (another region implicated in visual working memory of spatial information; Sprague et al., 2014, 2016) contained a representation throughout the memory delay period of the swapped item rather than the cued item. These results demonstrate the selection and active maintenance of an uncued (i.e., incorrect) memory representation that otherwise resembles accurate memory maintenance.

Methods

Participants

Twelve participants were recruited to the prescreening portion of the experiment, six of whom (3 female, 3 male) fit our swap rate criterion (between 20–40% of trials) and accepted an invitation to participate in the fMRI portion of the experiment. Participants were financially compensated at an hourly rate for the behavioral prescreening ($12) and fMRI ($20) portions of the experiment. The study was approved by the University of Texas Institutional Review Board.

Task

All participants completed two fMRI scanning sessions, each of which included an anatomical scan, four mapping task runs (not used in the current analysis), and a variable amount of memory task runs (runs were continued until the 2-hour time limit was reached or upon participant request). While our initial intention was to use the mapping task runs to build a perceptual spatial model that could be used to reconstruct memory representations (Sprague et al., 2014, 2016), this perception-to-memory generalization was unreliable, so this data was not used in the present analyses. Participants completed on average 208.3 memory trials (*min* = 200, *max* = 220). The memory task (Figure 1) was a version of a continuous response delayed estimation task (Ma et al., 2014; Wilken & Ma, 2004). To increase the likelihood of swap errors, all trials had a high set size (Bays, 2016) that exceeded typical working memory capacity (Oberauer et al., 2016). After initial trial fixation (0.1° radius, light gray; 500 ms), participants encoded an array of six colored discs (0.4° radius) presented 3.5° from center fixation (i.e., placed on an invisible ring with 3.5° radius). Locations were chosen uniformly between 0–359° (integers only), with the only restriction that all discs from a single trial be separated by at least 45° on the invisible ring. This minimum distance was implemented so that we could reconstruct each disc/location separately in the fMRI analysis. Each disc was colored with a unique hue from HSV color space, similarly sampled uniformly from integers 0–359° (note that saturation and value/brightness were set to 1 for all stimuli, but because the display was not gamma-corrected or color calibrated, luminance may have varied somewhat across stimuli). Minimum distance between hues of a single trial was 20° to minimize perceptual confusion. After encoding, a pre-cue delay (4300 ms) was presented before fixation changed color to match one of the six target discs, serving as a retro-cue (500 ms) (Griffin & Nobre, 2003) to inform participants which of the six disc locations to recall for the upcoming probe. Then a 12-second delay prefaced the memory probe, where participants were given six seconds to move a probe disc (0.4° radius, black outline) around the invisible ring. Probe start locations were also chosen uniformly between 0–359° (integers only). Participants used their right hand to move the probe disc around the wheel using the four buttons on a Cambridge Research Systems MRI-compatible "curved right" 4-button box. Buttons 1 and 2 (pushed with index and middle finger, respectively) were used to move the disc around the wheel clockwise (button 2) or counter-clockwise (button 1) at a coarse rate of 120° per second, while fine judgements could be made at 24° per second with buttons 3 (counter-clockwise) and 4 (clockwise) (ring and pinky finger, respectively). A subset of "drop" trials had a black retro-cue, indicating that there would be no memory test, but instead a perceptual task where participants would simply move the unfilled circle to a separate black disc placed randomly along the ring. After six seconds, wherever the probe disc was located was taken as their response. Immediately after the location probe, participants reported how confident they were about the accuracy of their response by choosing either "guess" (button 1), "low" (button 2), or "high" (button 3). Below each text option was a white ring that would fill in upon response selection. Participants were free to select and change their response for three seconds. After both location and confidence probes, location feedback was provided by showing the target disc overlaying the probe disc for one second. A 2-second inter-trial interval (ITI) ended every trial. The central fixation dot (0.1° radius, light gray) was present throughout the entire task except for ITIs, and participants were instructed to maintain strict fixation whenever it was on screen. During fMRI sessions, experimental displays were projected onto a screen using a VPixx Technologies PROPixx MRI projector. All experimental displays were coded and presented using PsychoPy software (Peirce, 2007, 2009) and presented on a gray background.

MRI acquisition parameters

Participants were scanned in a Siemens Skyra 3T scanner with a 32-channel head coil. Each scan session included a single high-resolution T1-weighted anatomical image (MEMPRAGE; FoV 256 mm, 256 × 256 matrix, 176 sagittal slices; TE=1.64/3.5/5.36/7.22 s). All functional scans were acquired using the same EPI sequence (TR=2 s; 76 slices; 3×3×3 mm voxel dimensions; 2x multiband factor).

MRI preprocessing

All anatomical and functional MRI data was preprocessed using fMRIprep 20.2.6 (Abraham et al., 2014; Esteban et al., 2019, 2020; Gorgolewski et al., 2011). Full processing details of the fMRIprep pipeline can be found in Esteban et al. (2019), but are briefly described here. Anatomical T1-weighted (T1w) scans were corrected for intensity non-uniformity (Tustison et al., 2010), skull-stripped, and tissue-segmented (Y. Zhang et al., 2001). A single T1w-reference map was created for each participant by registering their 2 T1w images (one from each scan session) using FreeSurfer (Reuter et al., 2010), and this image was used for brain surface reconstructions using FreeSurfer's recon-all (Dale et al., 1999). For each participant, a skull-stripped reference volume was created for all functional data, and then co-registered to the T1w-reference using FreeSurfer (Greve & Fischl, 2009). Finally, functional data was resampled onto native space correcting for head motion (Jenkinson et al., 2002).

Regions-of-interest

Region-of-interest masks were determined anatomically for each subject individually using a probabilistic retinotopic atlas (Wang et al., 2015) based on FreeSurfer surface reconstructions (see, MRI preprocessing). The early visual cortex (EVC) mask combines bilateral V1-V3a/b and hV4, whereas the intraparietal sulcus (IPS) mask combines bilateral IPS0–5. All individual EVC regions-of-interest provided qualitatively similar results, and thus were merged into a single mask.

Behavioral modeling

Each participant's responses were modeled as coming from a mixture distribution composed of three component distributions (Eq. 1). ϕ_κ represents the von Mises distribution with mean zero and concentration parameter κ. $\hat{\theta}$ denotes the location response on a given trial, where θ is the target location and $\{\theta_1, \theta_2, \ldots, \theta_m\}$ are the m non-target locations of the same trial (note m is constant at 5 for all trials in the current experiment). Maximum likelihood estimates of accurate (α), swap error (β), and guess (γ) response proportions were obtained for each subject using a non-linear optimization algorithm (Nelder & Mead, 1965). Data was aggregated across both scan sessions for each subject. This model is as described in (Bays et al., 2009) and the Analogue Report Toolbox was used for analysis (https://www.paulbays.com/toolbox/).

$$p(\hat{\theta}) = \alpha\phi_\kappa(\hat{\theta} - \theta) + \beta\frac{1}{m}\sum_{i}^{m}\phi_\kappa\left(\hat{\theta} - \theta_i\right) + \gamma\frac{1}{2\pi}$$

Equation (1)

For trial-level parameter estimates, we implemented the model as described in (Schneegans & Bays, 2016). Briefly here, this model takes, for each subject, the best-fitting mixture model parameters (described above) and uses Bayes theorem to derive the probability that a given trial came from each of the three mixture distributions (target, swap, and guess). The result is three probabilities, each of which corresponds to the probability of the trial coming from the target, swap, or guess distribution. Our post-hoc trial binning consisted of labeling each trial with the trial type of the highest probability. If the trial was labeled as a swap trial, we treated the nearest non-target memory item as the most likely swap candidate. The Analogue Report Toolbox was also used for this analysis.

fMRI modeling

Memory representations were evaluated using an inverted encoding model (Brouwer & Heeger, 2009; Sprague et al., 2014, 2015, 2018; Sprague & Serences, 2013). For each participant, using all fMRI runs concatenated across both sessions, we used a leave-one-run-out cross-validation procedure where we iteratively trained the encoding model on only the *accurate trials* of all but one run and built reconstructions on *all trials* of the held-out run. For the main analysis, to increase signal-to-noise ratio prior to modeling, we built a single memory representation for each trial by averaging the BOLD signal across the final three TRs (6 s) of the delay period of a given trial. For each cross-validated iteration of model training, linear regression was applied (Eq. 2) to all the training data (B_1: m voxels X n trials) to extract the weights (W: 36 channels X m voxels) – at each voxel – for each of 36 hypothetical spatial location basis functions that together comprise any given location (C_1: 36 channels X m voxels). The 36 derived location channel weights of each voxel were then applied (Eq. 3) to the held-out test data (B_2) to "reconstruct" a spatial location for each memory trial (C_2). For a more detailed account of the evolution of memory representations over the course of the trial, this procedure was also repeated at each TR in a subsequent analysis. Critically, for all analyses, each reconstruction was co-registered/rotated to a central location (3.5°, 0°) that allowed for averaging across trials. Altogether, we coregistered trials in four different ways: (1) on model-labeled accurate trials, we coregistered reconstructions with respect to the true memory target; on model-labeled swap trials, we coregistered reconstructions (2) with respect to the most likely swap candidate (i.e., nearest non-target spatial location) and (3) with respect to the true memory target (as a null condition); (4) on drop trials, we coregistered reconstructions with respect to a spatial location randomly chosen from the set of six locations encoded on that trial.

$$W = B_1 C_1^T \left(C_1 C_1^T\right)^{-1}$$

Equation (2)

$$C_2 = \left(W^T W\right)^{-1} W^T B_2$$

Equation (3)

To quantify the model reconstructions, we performed a resampling procedure across all trials and participants. For each condition, 1,000 times we took 240 random trials across all participants and averaged the co-registered two-dimensional reconstructions. We drew 240 samples with replacement to account for the mismatch of trial counts between accurate, swap, and drop trials; drop trials were frequently the lowest count at 40 trials per subject (240 total trials). Note that all effects were consistent across a variety of resampling procedures, such as instead sampling 180 trials within each trial type or sampling the first 40 trials within each condition per participant. Next, at each iteration, the average reconstruction was converted to a one-dimensional vector representation by extracting the values that encompassed the ring of stimulus representations on-screen (3.5±0.6° from center) and then converted to a representational fidelity measure (Sprague et al., 2014). Representational fidelity is the vector mean of the one-dimensional reconstruction (Eq. 4), where x_i is the reconstruction value at theta i for each of N = 220 ring segments. Thus, above-zero fidelity represents a peaked and "successful" reconstruction, whereas a fidelity of zero implies no memory representation. This procedure was repeated via bootstrap resampling across all trials and participants to generate group statistics, and p-values were generated by taking the fraction of resampled fidelity values above zero, and the fraction below zero, and multiplying the smallest fraction by two to account for a two-tailed test (Sprague et al., 2014, 2016). Similarly, resampled reconstruction fidelity values were compared between conditions by calculating the distribution of differences between each pair of resampled fidelity distributions and comparing these difference distributions to zero to calculate p-values. Finally, both these pairwise comparison p-values and the time course p-values (across all 10 TRs of the trial) were corrected for multiple comparisons within each analysis using the Benjamini-Hochberg false discovery rate method (Benjamini & Hochberg, 1995). The inverted encoding model analysis was implemented with custom Python code using the SciPy stack (Oliphant, 2007), Pandas (McKinney, 2010), and Nilearn (Abraham et al., 2014; Pedregosa et al., 2011), and Pingouin (v0.5.0) was used for statistical testing (Vallat, 2018).

$$F = \frac{1}{N}\sum_{i=1}^{N} x_i \cos i$$

Equation (4)

Results

Behavior

Mixture modeling revealed that, after prescreening, participants continued to show a roughly 30% swap rate (*M* = .30, *SEM* = .02) during their two fMRI sessions (Figure 2). Trial-level parameter estimates resulted in an average of 53 swap trials (*SEM* = 4) per participant (*min* = 36, *max* = 65). Raw response error tracked well with subjective confidence, suggesting a level of metacognition about their performance (Figure 3A). Error on reported high-confidence trials (*M* = 27°, *SEM* = 4.8°) was lower than that of low-confidence trials (*M* = 47°, *SEM* = 3.3°; *t*(5) = 4.53, *p* = .006, *CLES* = .94) and guess trials (*M* = 68°, *SEM* = 5.4°; *t*(5) = 5.29, *p* = .003, *CLES* = 1.00). Reported low-confidence trials also had lower error than guess trials (*t*(5) = 4.03, *p* = .01, *CLES* = .97).

To compare mixture model output with subjective confidence, we compared the proportion of guess, low confidence, and high confidence responses within both model-labeled accurate and swap trials. For each, we resampled 1,000 trials with replacement, each time calculating the proportion of trials that were subjectively reported as guess, low confidence, or high confidence. Resampled distributions within each confidence response were compared by resampling difference measures 1,000 times with replacement. Low confidence responses made up most of the trials, with equal proportions in accurate and swap trials (*M* = −.05, *CI* = [−.11, .02], *p* = .16). Participants reported more high confidence (*M* = .17, *CI* = [.12, .23], *p* < .001) and less guessing (*M* = −.12, *CI* = [−.18, −.07], *p* < .001) on accurate trials as compared to swap trials, suggesting a shift to more guesses, albeit with most responses (55%) on swap trials being reported as low confidence.

fMRI reconstructions

An inverted encoding model was applied to delay-period BOLD data in an attempt to reconstruct a variety of spatial locations in early visual cortex (EVC) and intraparietal sulcus

(IPS) across different trial types (Figure 4). On accurate trials, we found reliable memory representations during the delay period of the true memory target location in EVC ($M_{\text{fidelity}} = .05$, $CI = [.03, .06]$, $p < .001$) and IPS ($M_{\text{fidelity}} = .04$, $CI = [.03, .05]$, $p < .001$). On drop trials, when participants were not asked to remember anything, we were unable to reconstruct a randomly-chosen memory target location in either EVC ($M_{\text{fidelity}} = .01$, $CI = [-.01, .02]$, $p = .33$) or IPS ($M_{\text{fidelity}} = .01$, $CI = [-.01, .03]$, $p = .40$). Importantly, on swap trials, we were able to reconstruct representations of a non-target location (the swap target) in both EVC ($M_{\text{fidelity}} = .04$, $CI = [.02, .05]$, $p < .001$) and IPS ($M_{\text{fidelity}} = .03$, $CI = [.01, .04]$, $p < .001$). Moreover, the fidelity of the reconstructed representations did not differ between accurate and swap trials in either EVC ($M_{\text{fidelity\_accurate\_vs\_swap}} = .01$, $CI = [-.01, .03]$, $p_{\text{FDR}} = .27$) or IPS ($M_{\text{fidelity\_accurate\_vs\_swap}} = .02$, $CI = [0, .03]$, $p_{\text{FDR}} = .12$). In the EVC, the fidelities of both the accurate and swap reconstructions were significantly greater than the reconstruction fidelity on drop trials ($M_{\text{fidelity\_accurate\_vs\_drop}} = .04$, $CI = [.02, .06]$, $p_{\text{FDR}} < .001$; $M_{\text{fidelity\_swap\_vs\_drop}} = .03$, $CI = [.01, .05]$, $p_{\text{FDR}} = .003$). Finally, in the IPS, while the fidelity of accurate trial reconstructions was higher than drop trial reconstructions ($M_{\text{fidelity\_accurate\_vs\_drop}} = .03$, $CI = [.01, .06]$, $p_{\text{FDR}} = .012$), the fidelity of swap and drop trial reconstructions did not significantly differ ($M_{\text{fidelity\_swap\_vs\_drop}} = .02$, $CI = [0, .04]$, $p_{\text{FDR}} = .10$).

To gain more insight into how accurate and swap representations emerge over the course of the trial in the EVC, we performed the same inverted encoding model location reconstruction approach separately for each trial timepoint. On accurate trials, reliable memory reconstructions emerged about 6s after the retro-cue (note data is not shifted to account for hemodynamic lag; Figure 5). On swap trials, error representations followed a similar timecourse, with reliable representations beginning about 8 seconds after the retro-cue (Figure 6). As expected, on drop trials, reliable representations of randomly-chosen memory locations did not emerge at any point during the trial.

Discussion

The goal of the current study was to evaluate the neural correlate of working memory swap errors (the reporting of an incorrect item from a memory set). An increase in such memory errors is frequently associated with cognitive aging (Peich et al., 2013) and clinical memory decline associated with Alzheimer’s disease (Parra et al., 2009) and Parkinson’s disease (Schneegans & Bays, 2019; Zokaei et al., 2014), and therefore binding errors are considered a core feature that must be accounted for in any effective working memory model (Oberauer et al., 2018). Here, participants were scanned with fMRI while they performed a working memory task that induced a high rate of swap errors (~30%). After labeling trials post-hoc as accurate or swap errors using behavioral modeling (Schneegans & Bays, 2016), we used an inverted encoding model to test for the existence of mnemonic representations during the memory retention periods (Sprague et al., 2015). We reconstructed location-specific working memory representations from fMRI data and found that swap errors are preceded by active maintenance of that swapped item. The neural trace of an incorrect memory representation that appears otherwise intact prior to response is consistent with an explanation of swap errors as resulting from confusion between memory items that occurs prior to the time of a memory response.

We extended previous work showing that early visual areas and intraparietal sulcus contain mnemonic information for spatial items (Curtis, 2006; Sprague et al., 2014) by reconstructing memory representations of the cued location when participants responded accurately. The cued location could not be reconstructed on swap trials, but we discovered that an uncued location corresponding to the participant’s eventual behavioral report (the swap target) could be reconstructed throughout the delay period prior to the memory test. These results provide neural evidence for the existence of swap errors that are incorrect responses being maintained in working memory, rather than guesses being generated during the response period. Although inconclusive about the origin of swap errors, these data provide a useful starting point from which future studies can further disentangle the potential neural mechanisms leading to these active swap representations in working memory.

A recent critique of mixture modeling of working memory responses is that model-labeled swap errors may simply be (educated) guesses rather than true swaps in memory (Huang, 2020; Pratte, 2018). Our data contribute to this discussion in two ways. First, the active neural trace of swap candidates on swap trials – beginning about 8–10s before the behavioral response – suggests that swap errors do not exclusively arise from educated guesses at the time of test. Instead, the neural process giving rise to such errors occurs shortly after the retro-cue and results in sustained active representations that share the same representational format as accurate memory representations (evidenced by our ability to reconstruct swap representations with an inverted encoding model built on accurate trials). In the current task, the retro-cue serves as a response selection cue, and it is plausible that on swap trials participants made an educated guess to select and actively maintain an (incorrect) item throughout the delay period. Thus, we can’t definitively rule out an “educated guess” interpretation of swap trials (Pratte, 2018). However, these results suggest that if swaps do indeed reflect educated guesses, the resulting representations share the same representational format, a similar timecourse, and similar levels of fidelity as accurate memory representations. Second, participants in our study reported only a small percentage of their swap trials as subjective guesses (*M* = 24%, *SEM* = 5%). Previous work has shown that participants are metacognitively aware of their working memory performance (Rademaker et al., 2012; van den Berg et al., 2017), even if slightly overconfident at times (Adam & Vogel, 2017). In the current study, there was a higher proportion of subjective guesses on swap trials than on accurate trials. This is potentially consistent with the view that at least a subset of model-labeled swaps were in reality educated guesses (Pratte, 2018). However, our finding that both accurate and swap trials exhibit a high proportion of low-confidence responses challenges this view and suggests that the majority of model-labeled swap trials were responded to with non-zero confidence. A useful comparison would be to include the subjective confidence of model-labeled guesses, but we didn’t have enough guess trials for such a comparison. Finally, it is also possible that participants’ idiosyncratic response criteria make these confidence ratings difficult to interpret. For example, a participant could have reserved the “guess” response for completely random guesses, and responded with “low” confidence on trials in which they made informed, strategic guesses. If the strategic guesses were made immediately after the retro-cue, this could be consistent with our observed results. It would be informative for future work to provide more nuanced instructions to participants to allow for reports of both random and educated guesses. Furthermore, it would be beneficial to adopt a design that yields a higher proportion of model-labeled guesses (there were fewer than 5% in this study), in addition to a high proportion of swaps, to examine whether active representations are observed, for example, only on swap trials but not on trials with random guessing.

Why do swap errors occur? Unfortunately, with the current data, we are not able to identify the precise neural mechanism that produces swap errors. One possibility is that swap errors arise from binding errors between color and location feature information during the perception and/or encoding of the items into working memory. Some neural models propose that a separate neural population is dedicated to the binding of each individual feature, which have their own dedicated neural representations (Swan & Wyble, 2014). Other conjunctive coding models propose that a single neural population represents both features in tandem (Matthey et al., 2015; Oberauer & Lin, 2017; Schneegans & Bays, 2017). With conjunctive models, it’s possible that swap errors result directly from binding errors that occur due to the accumulation of neural noise in maintaining working memory items prior to the response probe (Bays, 2015; Schneegans & Bays, 2017). Other findings suggest that swap errors may not derive from binding errors between a cue feature and its response feature (e.g., a green cue and the 78° location), but rather from errors in interpreting the cue feature itself (e.g., mistaking the green cue as blue) (Bays, 2016; Emrich & Ferber, 2012; Rajsic & Wilson, 2014; Rerko et al., 2014). Importantly, experimental designs that use location as the cue feature and color as the response feature often provide results that differ from the less common color-as-cue and location-as-response designs (like ours). Swap errors are more common in the latter location-as-response designs (Rajsic & Wilson, 2014), potentially due to the unique status of location in visual working memory and/or its particular utility in an informed guessing strategy (Pratte, 2018). It would be informative for future work to examine whether active maintenance of non-target representations is also observed for stimulus features other than location.

We found reliable memorized location representations in both early visual cortex and in intraparietal sulcus, consistent with much previous work characterizing the wide range of brain regions involved in short-term memory maintenance (Christophel et al., 2017; Courtney et al., 1997; Ungerleider et al., 1998). Moreover, the redundant coding of memory representations along the visual hierarchy might be fundamental to resisting distraction (Lorenc et al., 2021; Lorenc & Sreenivasan, 2021). Working memory representations in early visual cortex retain precise visual representations of memoranda (Ester et al., 2013; Harrison & Tong, 2009), even in some cases amidst visual distraction (Rademaker et al., 2019). Parietal cortex might retain a working memory representation of the same item (Bettencourt & Xu, 2016; Xu & Chun, 2006), albeit a more abstract/categorical representation (Lee et al., 2013) that may be less susceptible to subtle perceptual interference (Lorenc et al., 2018). Furthermore, the role of parietal cortex in working memory is likely not restricted to item maintenance (Bays, 2018), as recent evidence suggests that the intraparietal sulcus serves an explicit role in aiding feature binding demands (Cai et al., 2020; Gosseries et al., 2018). In the current work, we held constant the context-binding demands, and thus could not perform a similar analysis.

The current results are limited by the small and specific sample size, although such sample sizes are common in related literature involving multiple scan sessions per participant (e.g., Rademaker et al., 2019; Sprague et al., 2014). All participants passed a pre-screening of high swap rates so that we would have enough swap trials for each participant for analyses. It remains a possibility that swap errors in low-swap participants are not actively maintained during the delay period. As highlighted above, our results are unable to definitively rule out an educated guess model where all swap errors are subjective guesses.

Acknowledgements.

This work was supported by the National Eye Institute of National Institutes of Health under Award Number R01EY028746 (JLP).

References

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

Adam KCS, & Vogel EK (2017). Confident failures: Lapses of working memory reveal a metacognitive blind spot. Attention, Perception, & Psychophysics, 79(5), 1506–1523. 10.3758/s13414-017-1331-8

Awh E, Anllo-Vento L, & Hillyard SA (2000). The Role of Spatial Selective Attention in Working Memory for Locations: Evidence from Event-Related Potentials. Journal of Cognitive Neuroscience, 12(5), 840–847. 10.1162/089892900562444 [PubMed: 11054925]

Bays PM (2015). Spikes not slots: Noise in neural populations limits working memory. Trends in Cognitive Sciences, 19(8), 431–438. 10.1016/j.tics.2015.06.004 [PubMed: 26160026]

Bays PM (2016). Evaluating and excluding swap errors in analogue tests of working memory. Scientific Reports, 6, 19203. 10.1038/srep19203 [PubMed: 26758902]

Bays PM (2018). Reassessing the Evidence for Capacity Limits in Neural Signals Related to Working Memory. Cerebral Cortex, 28(4). 10.1093/cercor/bhx351

Bays PM, Catalao RFG, & Husain M (2009). The precision of visual working memory is set by allocation of a shared resource. Journal of Vision, 9(10), 7–7. 10.1167/9.10.7 [PubMed: 19761322]

Benjamini Y, & Hochberg Y (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society: Series B (Methodological), 57(1), 289–300. 10.1111/j.2517-6161.1995.tb02031.x

Bettencourt KC, & Xu Y (2016). Decoding the content of visual short-term memory under distraction in occipital and parietal areas. Nature Neuroscience, 19(1), 150–157. 10.1038/nn.4174 [PubMed: 26595654]

Blacker KJ, & Courtney SM (2016). Distinct Neural Substrates for Maintaining Locations and Spatial Relations in Working Memory. Frontiers in Human Neuroscience, 10. 10.3389/fnhum.2016.00594

Brignani D, Bortoletto M, Miniussi C, & Maioli C (2010). The when and where of spatial storage in memory-guided saccades. NeuroImage, 52(4), 1611–1620. 10.1016/j.neuroimage.2010.05.039 [PubMed: 20493955]

Brouwer GJ, & Heeger DJ (2009). Decoding and Reconstructing Color from Responses in Human Visual Cortex. Journal of Neuroscience, 29(44), 13992–14003. 10.1523/JNEUROSCI.3577-09.2009 [PubMed: 19890009]

Cai Y, Fulvio JM, Yu Q, Sheldon AD, & Postle BR (2020). The Role of Location-Context Binding in Nonspatial Visual Working Memory. ENeuro, 7(6). 10.1523/ENEURO.0430-20.2020

Christophel TB, Klink PC, Spitzer B, Roelfsema PR, & Haynes J-D (2017). The Distributed Nature of Working Memory. Trends in Cognitive Sciences, 21(2), 111–124. 10.1016/j.tics.2016.12.007 [PubMed: 28063661]

Courtney SM, Ungerleider LG, Keil K, & Haxby JV (1997). Transient and sustained activity in a distributed neural system for human working memory. Nature, 386(6625), 386608a0. 10.1038/386608a0

Curtis CE (2006). Prefrontal and parietal contributions to spatial working memory. Neuroscience, 139(1), 173–180. 10.1016/j.neuroscience.2005.04.070 [PubMed: 16326021]

Dale AM, Fischl B, & Sereno MI (1999). Cortical Surface-Based Analysis: I. Segmentation and Surface Reconstruction. NeuroImage, 9(2), 179–194. 10.1006/nimg.1998.0395 [PubMed: 9931268]

Emrich SM, & Ferber S (2012). Competition increases binding errors in visual working memory. Journal of Vision, 12(4), 12–12. 10.1167/12.4.12

Esteban O, Ciric R, Finc K, Blair RW, Markiewicz CJ, Moodie CA, Kent JD, Goncalves M, DuPre E, Gomez DEP, Ye Z, Salo T, Valabregue R, Amlien IK, Liem F, Jacoby N, Stojić H, Cieslak M, Urchs S, … Gorgolewski KJ (2020). Analysis of task-based functional MRI data preprocessed with fMRIPrep. Nature Protocols, 15(7), 2186–2202. 10.1038/s41596-020-0327-3 [PubMed: 32514178]

Esteban O, Markiewicz CJ, Blair RW, Moodie CA, Isik AI, Erramuzpe A, Kent JD, Goncalves M, DuPre E, Snyder M, Oya H, Ghosh SS, Wright J, Durnez J, Poldrack RA, & Gorgolewski KJ (2019). fMRIPrep: A robust preprocessing pipeline for functional MRI. Nature Methods, 16(1), 111–116. 10.1038/s41592-018-0235-4 [PubMed: 30532080]

Ester EF, Anderson DE, Serences JT, & Awh E (2013). A Neural Measure of Precision in Visual Working Memory. Journal of Cognitive Neuroscience, 25(5), 754–761. 10.1162/jocn_a_00357 [PubMed: 23469889]

Golomb JD (2015). Divided spatial attention and feature-mixing errors. Attention, Perception, & Psychophysics, 77(8), 2562–2569. 10.3758/s13414-015-0951-0

Golomb JD, L'Heureux ZE, & Kanwisher N (2014). Feature-binding errors after eye movements and shifts of attention. Psychological Science, 25(5), 1067–1078. 10.1177/0956797614522068 [PubMed: 24647672]

Gorgolewski K, Burns CD, Madison C, Clark D, Halchenko YO, Waskom ML, & Ghosh SS (2011). Nipype: A Flexible, Lightweight and Extensible Neuroimaging Data Processing Framework in Python. Frontiers in Neuroinformatics, 5. 10.3389/fninf.2011.00013

Gosseries O, Yu Q, LaRocque JJ, Starrett MJ, Rose NS, Cowan N, & Postle BR (2018). Parietal-Occipital Interactions Underlying Control- and Representation-Related Processes in Working Memory for Nonspatial Visual Features. Journal of Neuroscience, 38(18), 4357–4366. 10.1523/JNEUROSCI.2747-17.2018 [PubMed: 29636395]

Greve DN, & Fischl B (2009). Accurate and robust brain image alignment using boundary-based registration. NeuroImage, 48(1), 63–72. 10.1016/j.neuroimage.2009.06.060 [PubMed: 19573611]

Griffin IC, & Nobre AC (2003). Orienting attention to locations in internal representations. Journal of Cognitive Neuroscience, 15(8), 1176–1194. 10.1162/089892903322598139 [PubMed: 14709235]

Harrison SA, & Tong F (2009). Decoding reveals the contents of visual working memory in early visual areas. Nature, 458(7238), nature07832. 10.1038/nature07832

Huang L (2020). Distinguishing target biases and strategic guesses in visual working memory. Attention, Perception, & Psychophysics, 82, 1258–1270. 10.3758/s13414-019-01913-2

Jenkinson M, Bannister P, Brady M, & Smith S (2002). Improved Optimization for the Robust and Accurate Linear Registration and Motion Correction of Brain Images. NeuroImage, 17(2), 825–841. 10.1006/nimg.2002.1132 [PubMed: 12377157]

Kiyonaga A, Scimeca JM, Bliss DP, & Whitney D (2017). Serial Dependence across Perception, Attention, and Memory. Trends in Cognitive Sciences, 21(7), 493–497. 10.1016/j.tics.2017.04.011 [PubMed: 28549826]

Lee S-H, Kravitz DJ, & Baker CI (2013). Goal-dependent dissociation of visual and prefrontal cortices during working memory. Nature Neuroscience, 16(8), 997–999. 10.1038/nn.3452 [PubMed: 23817547]

Lorenc ES, Mallett R, & Lewis-Peacock JA (2021). Distraction in Visual Working Memory: Resistance is Not Futile. Trends in Cognitive Sciences. 10.1016/j.tics.2020.12.004

Lorenc ES, & Sreenivasan KK (2021). Reframing the debate: The distributed systems view of working memory. Visual Cognition, 29(7), 416–424. 10.1080/13506285.2021.1899091

Lorenc ES, Sreenivasan KK, Nee DE, Vandenbroucke ARE, & D'Esposito M (2018). Flexible coding of visual working memory representations during distraction. Journal of Neuroscience, 38(23), 5267–5276. 10.1523/JNEUROSCI.3061-17.2018 [PubMed: 29739867]

Luck SJ, & Vogel EK (1997). The capacity of visual working memory for features and conjunctions. Nature, 390(6657), 279–281. 10.1038/36846 [PubMed: 9384378]

Luck SJ, & Vogel EK (2013). Visual working memory capacity: From psychophysics and neurobiology to individual differences. Trends in Cognitive Sciences, 17(8), 391–400. 10.1016/j.tics.2013.06.006 [PubMed: 23850263]

Ma WJ, Husain M, & Bays PM (2014). Changing concepts of working memory. Nature Neuroscience, 17(3), nn.3655. 10.1038/nn.3655

Matthey L, Bays PM, & Dayan P (2015). A Probabilistic Palimpsest Model of Visual Short-term Memory. PLOS Computational Biology, 11(1), e1004003. 10.1371/journal.pcbi.1004003 [PubMed: 25611204]

McKinney W (2010). Data structures for statistical computing in python. Proceedings of the 9th Python in Science Conference, 445, 51–56. http://conference.scipy.org/proceedings/scipy2010/mckinney.html

Miller GA, Galanter E, & Pribram KH (1960). Plans and the structure of behavior. Henry Holt and Co. 10.1037/10039-000

Munneke J, Heslenfeld DJ, & Theeuwes J (2010). Spatial working memory effects in early visual cortex. Brain and Cognition, 72(3), 368–377. 10.1016/j.bandc.2009.11.001 [PubMed: 19962813]

Myers NE, Stokes MG, & Nobre AC (2017). Prioritizing Information during Working Memory: Beyond Sustained Internal Attention. Trends in Cognitive Sciences, 21(6), 449–461. 10.1016/j.tics.2017.03.010 [PubMed: 28454719]

Nelder JA, & Mead R (1965). A Simplex Method for Function Minimization. The Computer Journal, 7(4), 308–313. 10.1093/comjnl/7.4.308

Oberauer K, Farrell S, Jarrold C, & Lewandowsky S (2016). What limits working memory capacity? Psychological Bulletin, 142(7), 758–799. 10.1037/bul0000046 [PubMed: 26950009]

Oberauer K, Lewandowsky S, Awh E, Brown GDA, Conway A, Cowan N, Donkin C, Farrell S, Hitch GJ, Hurlstone MJ, Ma WJ, Morey CC, Nee DE, Schweppe J, Vergauwe E, & Ward G (2018). Benchmarks for models of short-term and working memory. Psychological Bulletin, 144(9), 885–958. 10.1037/bul0000153 [PubMed: 30148379]

Oberauer K, & Lin H-Y (2017). An interference model of visual working memory. Psychological Review, 124(1), 21–59. 10.1037/rev0000044 [PubMed: 27869455]

Oliphant TE (2007). Python for Scientific Computing. Computing in Science Engineering, 9(3), 10–20. 10.1109/MCSE.2007.58

Parra MA, Abrahams S, Fabi K, Logie R, Luzzi S, & Sala SD (2009). Short-term memory binding deficits in Alzheimer's disease. Brain, 132(4), 1057–1066. 10.1093/brain/awp036 [PubMed: 19293236]

Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, Vanderplas J, Passos A, Cournapeau D, Brucher M, Perrot M, & Duchesnay É (2011). Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research, 12(Oct), 2825–2830.

Peich M-C, Husain M, & Bays PM (2013). Age-related decline of precision and binding in visual working memory. Psychology and Aging, 28(3), 729–743. 10.1037/a0033236 [PubMed: 23978008]

Peirce JW (2007). PsychoPy—Psychophysics software in Python. Journal of Neuroscience Methods, 162(1), 8–13. 10.1016/j.jneumeth.2006.11.017 [PubMed: 17254636]

Peirce JW (2009). Generating stimuli for neuroscience using PsychoPy. Frontiers in Neuroinformatics, 2. 10.3389/neuro.11.010.2008

Pertzov Y, Dong MY, Peich M-C, & Husain M (2012). Forgetting What Was Where: The Fragility of Object-Location Binding. PLOS ONE, 7(10), e48214. 10.1371/journal.pone.0048214 [PubMed: 23118956]

Peters B, Kaiser J, Rahm B, & Bledowski C (2015). Activity in Human Visual and Parietal Cortex Reveals Object-Based Attention in Working Memory. Journal of Neuroscience, 35(8), 3360–3369. 10.1523/JNEUROSCI.3795-14.2015 [PubMed: 25716836]

Pratte MS (2018). Swap errors in spatial working memory are guesses. Psychonomic Bulletin & Review. 10.3758/s13423-018-1524-8

Pratte MS, & Tong F (2014). Spatial specificity of working memory representations in the early visual cortex. Journal of Vision, 14(3), 22–22. 10.1167/14.3.22

Rademaker RL, Chunharas C, & Serences JT (2019). Coexisting representations of sensory and mnemonic information in human visual cortex. Nature Neuroscience, 22(8), 1336. 10.1038/s41593-019-0428-x [PubMed: 31263205]

Rademaker RL, Tredway CH, & Tong F (2012). Introspective judgments predict the precision and likelihood of successful maintenance of visual working memory. Journal of Vision, 12(13), 21–21. 10.1167/12.13.21

Rajsic J, & Wilson DE (2014). Asymmetrical access to color and location in visual working memory. Attention, Perception, & Psychophysics, 76(7), 1902–1913. 10.3758/s13414-014-0723-2

Rerko L, Oberauer K, & Lin H-Y (2014). Spatial Transposition Gradients in Visual Working Memory. Quarterly Journal of Experimental Psychology, 67(1), 3–15. 10.1080/17470218.2013.789543

Reuter M, Rosas HD, & Fischl B (2010). Highly accurate inverse consistent registration: A robust approach. NeuroImage, 53(4), 1181–1196. 10.1016/j.neuroimage.2010.07.020 [PubMed: 20637289]

Schneegans S, & Bays PM (2016). No fixed item limit in visuospatial working memory. Cortex, 83, 181–193. 10.1016/j.cortex.2016.07.021 [PubMed: 27565636]

Schneegans S, & Bays PM (2017). Neural Architecture for Feature Binding in Visual Working Memory. The Journal of Neuroscience, 37(14), 3913. 10.1523/JNEUROSCI.3493-16.2017 [PubMed: 28270569]

Schneegans S, & Bays PM (2019). New perspectives on binding in visual working memory. British Journal of Psychology, 110(2), 207–244. 10.1111/bjop.12345 [PubMed: 30295318]

Schurgin MW, Wixted JT, & Brady TF (2020). Psychophysical scaling reveals a unified theory of visual memory strength. Nature Human Behaviour, 4(11), 1156–1172. 10.1038/s41562-020-00938-0

Sprague TC, Adam KCS, Foster JJ, Rahmati M, Sutterer DW, & Vo VA (2018). Inverted Encoding Models Assay Population-Level Stimulus Representations, Not Single-Unit Neural Tuning. ENeuro, 5(3), 1–5. 10.1523/ENEURO.0098-18.2018

Sprague TC, Ester EF, & Serences JT (2014). Reconstructions of Information in Visual Spatial Working Memory Degrade with Memory Load. Current Biology, 24(18), 2174–2180. 10.1016/j.cub.2014.07.066 [PubMed: 25201683]

Sprague TC, Ester EF, & Serences JT (2016). Restoring Latent Visual Working Memory Representations in Human Cortex. Neuron, 91(3), 694–707. 10.1016/j.neuron.2016.07.006 [PubMed: 27497224]

Sprague TC, Saproo S, & Serences JT (2015). Visual attention mitigates information loss in small- and large-scale neural codes. Trends in Cognitive Sciences, 19(4), 215–226. 10.1016/j.tics.2015.02.005 [PubMed: 25769502]

Sprague TC, & Serences JT (2013). Attention modulates spatial priority maps in the human occipital, parietal and frontal cortices. Nature Neuroscience, 16(12), nn.3574. 10.1038/nn.3574

Swan G, & Wyble B (2014). The binding pool: A model of shared neural resources for distinct items in visual working memory. Attention, Perception, & Psychophysics, 76(7), 2136–2157. 10.3758/s13414-014-0633-3

Treisman A (1996). The binding problem. Current Opinion in Neurobiology, 6(2), 171–178. 10.1016/S0959-4388(96)80070-5 [PubMed: 8725958]

Treisman A, & Schmidt H (1982). Illusory conjunctions in the perception of objects. Cognitive Psychology, 14(1), 107–141. 10.1016/0010-0285(82)90006-8 [PubMed: 7053925]

Tustison NJ, Avants BB, Cook PA, Zheng Y, Egan A, Yushkevich PA, & Gee JC (2010). N4ITK: Improved N3 Bias Correction. IEEE Transactions on Medical Imaging, 29(6), 1310–1320. 10.1109/TMI.2010.2046908 [PubMed: 20378467]

Ungerleider LG, Courtney SM, & Haxby JV (1998). A neural system for human visual working memory. Proceedings of the National Academy of Sciences, 95(3), 883–890. 10.1073/pnas.95.3.883

Vallat R (2018). Pingouin: Statistics in Python. Journal of Open Source Software, 3(31), 1026. 10.21105/joss.01026

van den Berg R, Awh E, & Ma WJ (2014). Factorial comparison of working memory models. Psychological Review, 121(1), 124–149. 10.1037/a0035234 [PubMed: 24490791]

van den Berg R, Yoo AH, & Ma WJ (2017). Fechner's law in metacognition: A quantitative model of visual working memory confidence. Psychological Review, 124(2), 197–214. 10.1037/rev0000060 [PubMed: 28221087]

Wang L, Mruczek REB, Arcaro MJ, & Kastner S (2015). Probabilistic Maps of Visual Topography in Human Cortex. Cerebral Cortex, 25(10), 3911–3931. 10.1093/cercor/bhu277 [PubMed: 25452571]

Wheeler ME, & Treisman AM (2002). Binding in short-term visual memory. Journal of Experimental Psychology: General, 131(1), 48–64. 10.1037/0096-3445.131.1.48 [PubMed: 11900102]

Wilken P, & Ma WJ (2004). A detection theory account of change detection. Journal of Vision, 4(12), 11–11. 10.1167/4.12.11

Xu Y, & Chun MM (2006). Dissociable neural mechanisms supporting visual short-term memory for objects. Nature, 440(7080), 91–95. 10.1038/nature04262 [PubMed: 16382240]

Zhang W, & Luck SJ (2008). Discrete fixed-resolution representations in visual working memory. Nature, 453(7192), 233–235. 10.1038/nature06860 [PubMed: 18385672]

Zhang Y, Brady M, & Smith S (2001). Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm. IEEE Transactions on Medical Imaging, 20(1), 45–57. 10.1109/42.906424 [PubMed: 11293691]

Zokaei N, McNeill A, Proukakis C, Beavan M, Jarman P, Korlipara P, Hughes D, Mehta A, Hu MTM, Schapira AHV, & Husain M (2014). Visual short-term memory deficits associated with GBA mutation and Parkinson's disease. Brain, 137(8), 2303–2311. 10.1093/brain/awu143 [PubMed: 24919969]

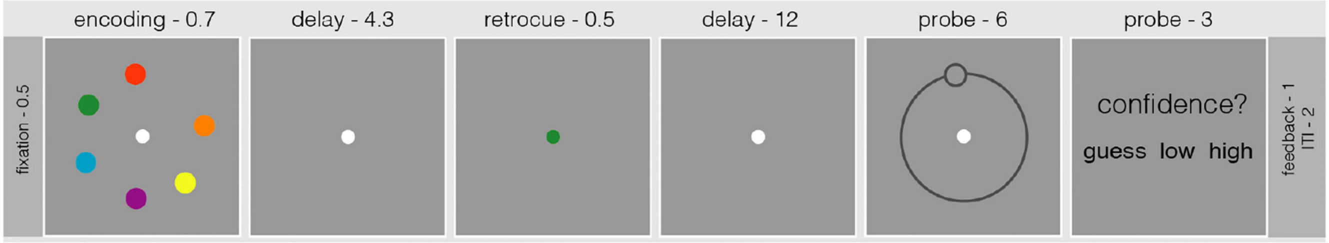

Figure 1. Memory task.

Participants encoded 6 colored discs, and in the middle of a lengthy delay were informed (via retro-cue) which location they would be tested on. All times in seconds.

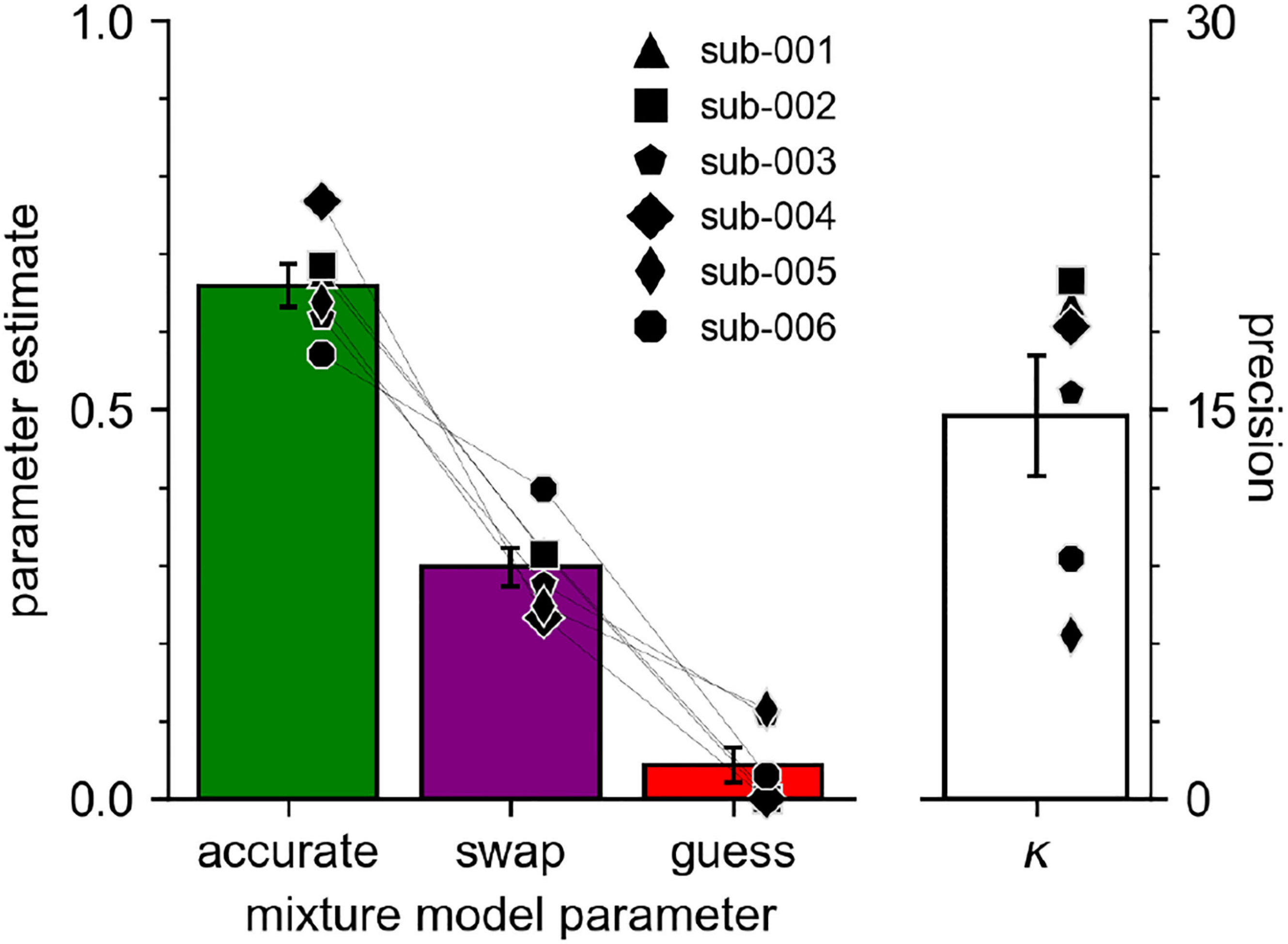

Figure 2. Mixture modeling results.

Each participant shows a roughly 30% swap rate across the 2 fMRI sessions.

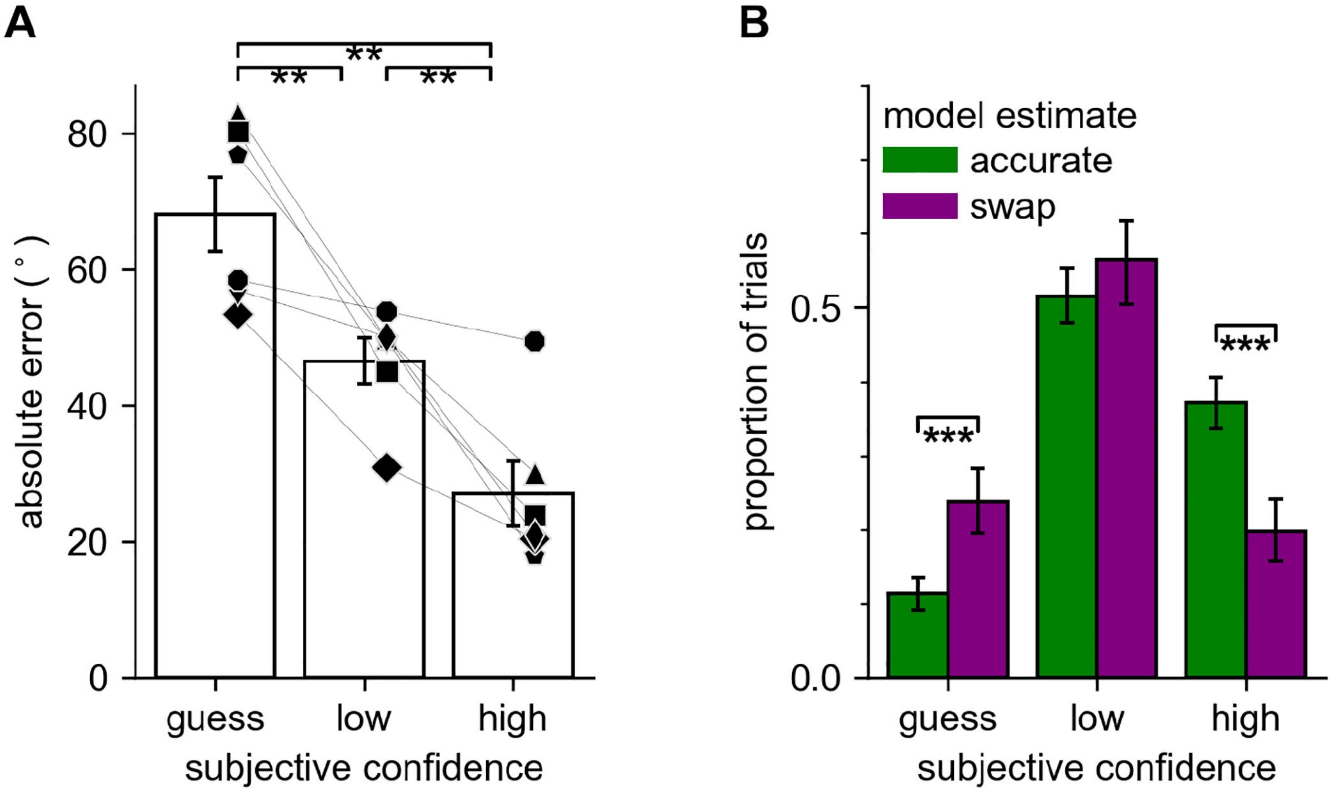

Figure 3. Confidence results.

A. Raw response error tracks well with reported confidence. B. Mixture model parameter proportions within each of model-labeled accurate and model-labeled swap trials. Participants report low confidence at similar rates for accurate and swap trials, whereas confidence shifts from high confidence to guess for swap trials. ** p < .01, *** p < .001

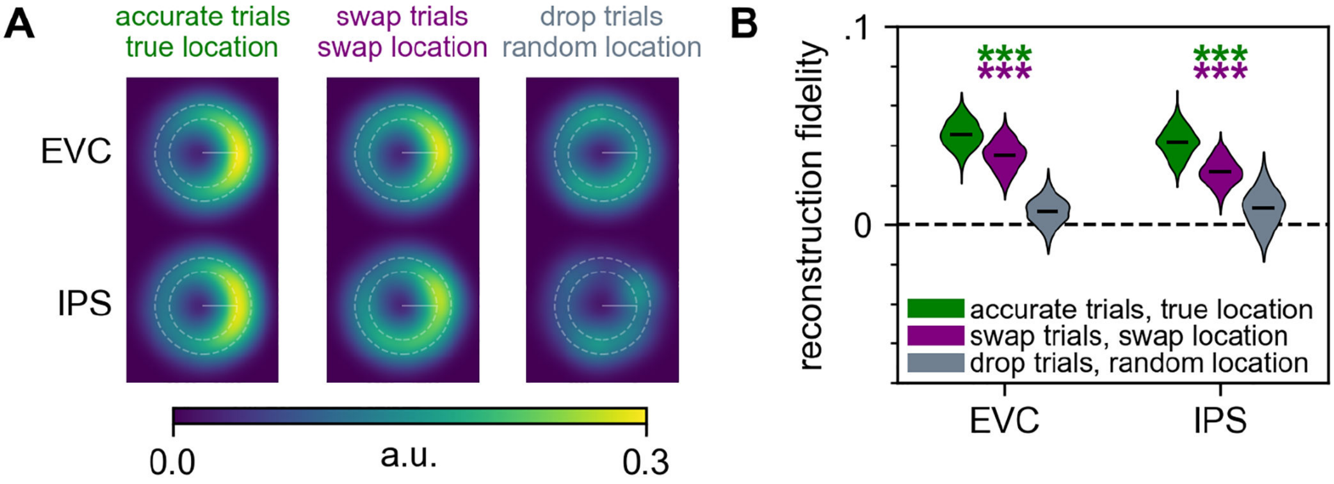

Figure 4. Memory delay reconstructions.

A. Raw two-dimensional reconstructions averaged across all trials and participants, in early visual cortex (EVC) and intraparietal cortex (IPS). B. The same reconstructions translated to representational fidelity measure through resampling procedure. On drop trials, reconstructions were recentered to a randomly chosen memory item location from that trial. Violins represent resampled mean distributions and black bars are means. *** p < .001 tested against zero

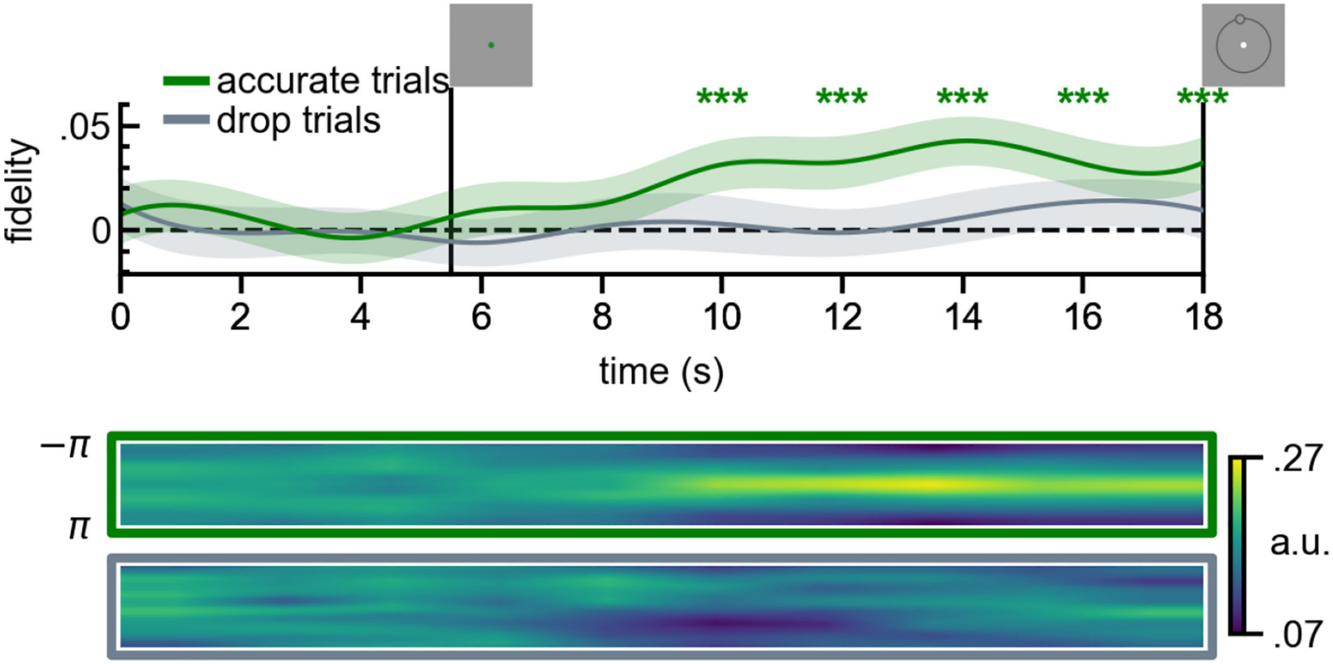

Figure 5. Accurate trial reconstruction timecourses in early visual cortex.

Top panel shows the timecourse of reconstructions across a trial, where the IEM is trained and tested through cross-validation at each timepoint/TR. Separate reconstructions were created for accurate trials (green) and drop trials (gray), and coregistered to the true memory target (accurate trials) or to a randomly-chosen target from the encoding array (drop trials). Bottom panel shows the resampled one-dimensional reconstructions, which is a middle step in analysis between the raw two-dimensional reconstructions and the final fidelity measure. * p < .05, ** p < .01, *** p < .001, FDR-corrected

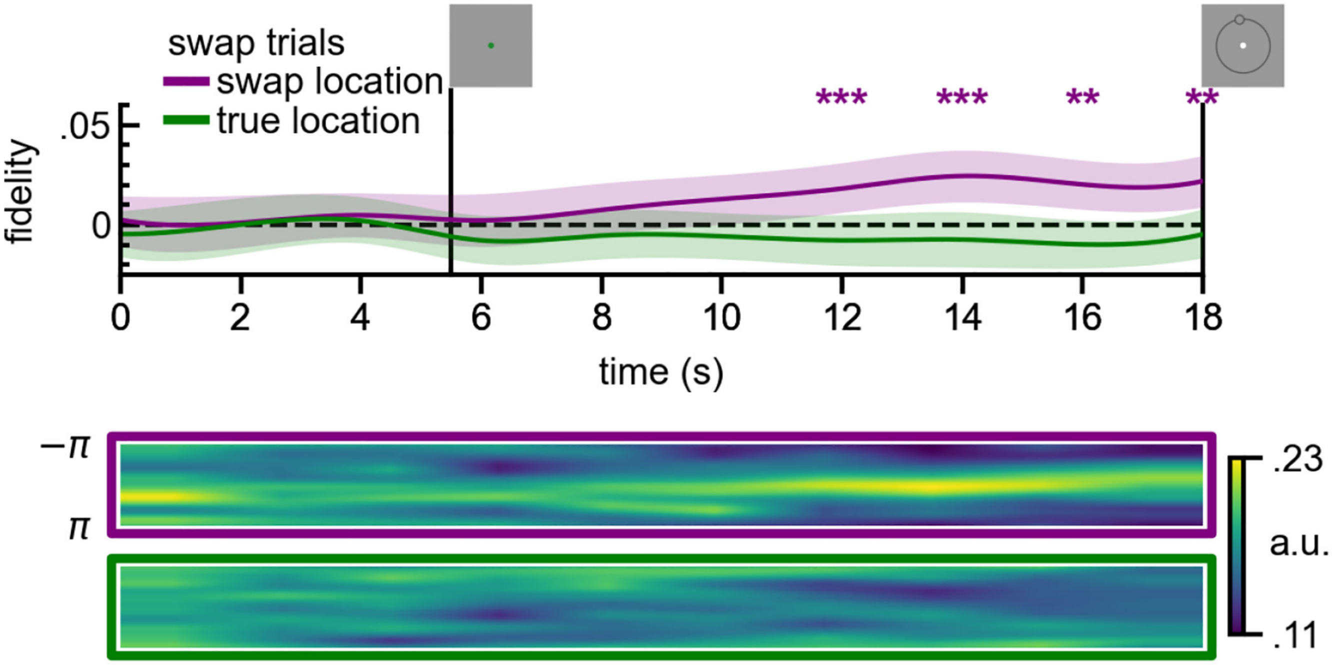

Figure 6. Swap trial reconstruction timecourses in early visual cortex.

Top panel shows the timecourse of reconstructions across a swap trial, where IEM is trained and test through cross-validation at each timepoint/TR. Data is tested on swap trials only, coregistering reconstructions to either the most-likely swap location (purple) or the true memory location (green). Bottom panel shows the resampled one-dimensional reconstructions, which is a middle step in analysis between the raw two-dimensional reconstructions and the final fidelity measure. * p < .05, ** p < .01, *** p < .001, FDR-corrected