eNeuro

Research Article: New Research | Cognition and Behavior

Cognitive flexibility improves memory for delayed intentions

https://doi.org/10.1523/ENEURO.0250-19.2019

Cite as: eNeuro 2019; 10.1523/ENEURO.0250-19.2019

Received: 27 June 2019
Revised: 7 September 2019
Accepted: 26 September 2019

Title: Cognitive flexibility improves memory for delayed intentions
Abbreviated Title: Cognitive flexibility improves delayed memory

Authors: Seth R. Koslov[1], Arjun Mukerji[2], Katlyn R. Hedgpeth[1], and Jarrod A. Lewis-Peacock[1]
Affiliations: [1]Department of Psychology, Institute for Neuroscience, University of Texas, Austin, Texas, [2]Helen Wills Neuroscience Institute, University of California (UC), Berkeley, California
Contributions: S.R.K., A.M., and J.L.P. conceived the study. S.R.K. and J.L.P. designed the experiment and wrote the paper. S.R.K. and K.H. collected and analyzed data under the supervision of J.L.P.
Corresponding Author: Jarrod A. Lewis-Peacock, email: jalewpea@utexas.edu

Abstract

The ability to delay execution of a goal until the appropriate time, prospective memory (PM), can be supported by two different cognitive control strategies: *proactive control* involving working memory maintenance of the goal and active monitoring of the environment, or *reactive control* relying on timely retrieval of goal information from episodic memory. Certain situations tend to favor each strategy, but the manner in which individuals adjust their strategy in response to changes in the environment is unknown. Across two experiments, human participants performed a delayed-recognition PM task embedded in an ongoing visual search task that fluctuated in difficulty. Control strategy was identified from moment to moment using reaction time costs and fMRI measures of goal maintenance. We found that people fluidly modified control strategies in accordance with changes in task demands (e.g., shifting towards proactive control when task difficulty decreased). This cognitive flexibility proved adaptive as it was associated with improved PM performance.

Significance Statement

Adapting to changes in the environment is important for achieving immediate goals, and it is also essential for remembering to perform future intentions. Using brain imaging and behavioral measures of cognitive control, we discovered that people fluidly shift between proactive and reactive control strategies, from moment to moment, in accordance with changes in ongoing task demands in order to successfully fulfill future intentions. These flexible shifts in control strategy were associated with better memory for delayed intentions, demonstrating that fine-grained control of attention and memory resources serves an adaptive role for remembering to carry out future plans.

Introduction

Life is busy and keeping track of what we are doing and what we intend to do can be challenging. Cognitive control describes the set of processes by which we are able to maintain and connect goals to actions and to subsequently filter out irrelevant distractors in accordance with these goals (Gratton, Cooper, Fabiani, Carter, & Karayanidis, 2018). Juggling goals in spite of interruptions, an ability known as prospective memory (PM), is a ubiquitous part of everyday life (Dismukes, 2012), constituting upwards of 50-80% of our daily memory problems (Crovitz & Daniel, 1984; Kliegel & Martin, 2003). The multiprocess theory of PM (McDaniel & Einstein, 2000; Einstein & McDaniel, 2005) describes two dissociable strategies that can be used: *proactive control* and *reactive control* (see also: Braver, 2012). Proactive control relies on working memory to remember the goal and external attention to monitor the environment for cues to act (Brewer, Knight, Marsh, & Unsworth, 2010; Guynn, 2003; McDaniel & Einstein, 2000; Smith, 2003). Reactive control relies on episodic memory to store the goal and salient cues from the environment to trigger its timely retrieval (McDaniel & Einstein, 2007, Einstein & McDaniel, 2010; Marklund & Persson, 2012).

These strategies have been shown to have distinct behavioral and neural profiles. Proactive control is cognitively demanding (Braver, Gray, & Burgess, 2007; Cohen, 2008) and interferes with ongoing processing (Smith, 2003; Smith, Hunt, McVay, & McConnell, 2007), whereas reactive control relies less on working memory processing and can succeed without any observable interference costs (Harrison & Einstein, 2010; Harrison, Mullet, Whiffen, Ousterhout, & Einstein, 2014; Knight et al., 2011; Rummel & Meiser, 2013; Scullin, McDaniel, & Einstein, 2010; Scullin, McDaniel,

Shelton, & Lee, 2010). Dissociable neural correlates have also been identified for these strategies (Beck, Ruge, Walser, & Goschke, 2014; Cona, Bisiacchi, & Moscovitch, 2014; Lewis-Peacock, Cohen, & Norman, 2016; McDaniel, LaMontagne, Beck, Scullin, & Braver, 2013; Reynolds, West, & Braver, 2009).

Preparatory processes involved in proactive control may benefit PM, but also place high costs on working memory and attentional capacities. In low-demand environments, controlled attentional processes can be successfully allocated to maintain goal information in working memory and to strategically monitor the environment for the right time and place to act (Braver et al., 2007; West, Bowry, & Krompinger, 2006; West, McNerney, & Travers, 2007). However, in high-demand environments, it is more efficient to offload the PM intention to the reactive control system and redirect cognitive resources towards more immediate demands (Braver, 2012). While reactive control is less cognitively demanding, it is more susceptible to proactive interference and more vulnerable to lapses in attention to goal-relevant events in the environment (Braver et al., 2007; Scullin, Bugg, & McDaniel, 2012). It is therefore important to select a control strategy best suited to the current situation in order to reduce the risk that prospective intentions interfere with more urgent demands, and also to reduce the risk that those intentions go unfulfilled.

One major area of research in PM has been to explain how individuals choose strategies in response to variable environmental demands (Anderson, McDaniel, & Einstein, 2017). The dynamic multiprocess framework (DMPV: Scullin, McDaniel, & Shelton, 2013; Shelton & Scullin, 2017), proposes that people have a flexible choice between proactive control and reactive control that primarily depends on the contextual

likelihood of a PM event. According to this model, individuals will rely on reactive control when the probability of a PM event is low, and then abruptly "switch on" proactive control when an environmental cue signals an increased likelihood of a PM event (see also: Ball, Brewer, Loft, & Bowden, 2015; Cohen, Gordon, Jaudas, Hefer, & Dreisbach, 2017; Kuhlmann & Rummel, 2014). However, previous work has primarily relied on blocked experiment designs or taken an all-or-none approach to measuring PM strategy. The environment may not always change abruptly, but instead may fluctuate more gradually from moment to moment. Correspondingly, PM strategy may fluctuate gradually between varying degrees of proactive and reactive control in response to the changing demands in the environment. The present study sought to test this hypothesis, and to evaluate the impact of strategy flexibility on PM performance.

In this study we tested the dynamics of PM strategy use by evaluating how individuals adjust their PM strategies when cognitive demands subtly increase or decrease over time, and in turn how this impacts prospective remembering. We hypothesized that adapting one's control strategy to better align with the cognitive load caused by competing demands should improve prospective remembering. Participants made a delayed target detection with pictures of faces and scenes (the PM task) while also performing an ongoing visual-search task with oriented arrows. The cognitive demands of the ongoing task were manipulated by subtly, but monotonically, adjusting task difficulty every couple of seconds. We linked time-sensitive behavioral measures of *PM strategy shifts* and multivariate neural measures of *PM intention processing* to memory performance on a trial-by-trial basis.

Materials and Methods

Human subjects were recruited at a location which will be identified if the article is published, and the experiments were carried out in a manner consistent with the approval of the internal review board of our institution. For experiment 1, 55 participants were recruited. Five of these participants were excluded due to below chance (<50%) performance on the ongoing task across the experiment. For experiment 2, we recruited 30 participants, two of which were excluded for excessive movement during scanning that led to MR images failing quality control. Data analyses were performed on the remaining 78 participants (28 neural sample, 17 F, mean age = 21.8; 50 behavioral sample, 31 F, mean age = 19.2).

Participants performed an ongoing visual search task with an embedded PM task (**Figure 1**). For the behavioral sample, participants performed six blocks of the PM task. For the neural sample, participants first completed one PM practice block outside of the scanner to familiarize themselves with the dual-task paradigm. During the practice block, face and scene stimuli were replaced with tools and vehicles so that participants did not familiarize themselves with the main task stimuli, but all other task configurations remained the same. Following completion of the practice session, participants were asked to explain the task, and the researcher answered any lingering questions before placing the subject in the MR scanner.

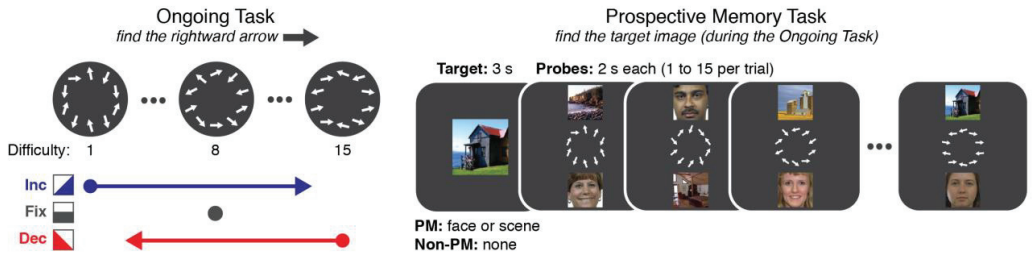

Figure 1. Task design. *Left*: Ongoing task difficulty could increase or decrease every 2s across a trial or remain fixed at the middle difficulty level (level 8 of 15). We validated the relationship between difficulty levels in a pilot study, finding that as difficulty increased, reaction time increased and accuracy decreased. For more information on pilot study 1 see **extended data Figure 1-1**. We further validated that the ongoing task could impact prospective memory strategy use in a second pilot study, where we found PM cost was significantly higher at an easy difficulty (level 4) than at a harder difficulty (level12). For more information on pilot study 2, see **extended data Figure 1-2**. *Right*: In the dual-task PM experiment, participants identified the reappearance of a PM target while concurrently performing the ongoing task.

Ongoing Task Details

The ongoing task was a visual-search task ("OG task") where participants searched for a specific target arrow on a circular array of oriented arrows (**Figure 1A**). We chose this as the OG task due to the ability to systematically and parametrically manipulate task difficulty by adjusting distractor parameters along a continuum (Sobel, Gerrie, Poole, & Kane, 2007; Kiyonaga, Dowd, & Egner, 2017). The ongoing task target arrow was always a rightward facing horizontal arrow (⇨). The target arrow was present on only a randomly selected half of the trials, located in one of 10 semi-randomly selected possible locations around the circle. Participants were instructed to search for the target on each display and use their right hand to press "1" for present and "2" for absent (for the MRI sample, that was buttons 1 and 2 on the response box). Target arrow location was counterbalanced between the top and bottom half of the screen. On

"present" trials, nine non-target (distractor) arrows appeared in set positions around the circular array (10 on "absent" trials), oriented within some distribution of angles determined by the current task difficulty setting.

OG task difficulty was manipulated on each probe by adjusting two parameters that determined the orientation of the distractor arrows: their minimum similarity to the target, and their similarity to other distractors. For distractor-to-target similarity, a minimum angular distance for distractors from the target (i.e., horizontal of 0°) was set to either 5°, 15°, 25°, 45°, 65°, or 75°. For distractor-to-distractor similarity, the maximum variance from the minimum angular distance was set to either 10°, 20°, or 40°. The factorial combination of these parameters (excluding any combination where minimum distance plus variance could exceed the 90° vertical plane) created 15 difficulty conditions that gradually and monotonically changed in difficulty. On every search display, each distractor arrow had a 50% chance of being randomly flipped across the horizontal plane and a 50% chance of being flipped across the vertical plane, so that distractor arrows could vary from horizontal across either 5°-175° or 185°-355°. In order to minimize uncontrolled pop-out effects, we ensured that no distractor arrow in the array was within 5° of an arrow that appeared in the same location on the circle during the previous display. This manipulation ensured that each visual-search array was a completely randomized new array of possible targets and distractors. We validated the relationship between difficulty levels in a pilot study, finding that as difficulty increased, reaction time increased $F(13,182)=39.53$, $p<.001$) and accuracy decreased $F(13,182)=74.89$, $p<.001$). For more information on pilot study 1 see **extended data Figure 1-1**.

For the behavioral sample, participants sat approximately 18 in away from the screen, and all 10 arrows, which were .64° by .22° in shape, were 3.18° away from the center of the screen. For the MRI sample, stimuli were projected on to a screen which participants viewed through a mirror placed over the headcoil. Participants laid face up, looking through a mirror at a projection screen that was approximately 136 cm away from the mirror. In order to keep the stimulus proportions relative to the field of view the same across experiments, projected arrows were 30° x 11°, and placed 1.39° away from the center of the screen.

PM Task Stimuli

Colored images of unfamiliar faces and unfamiliar scenes were gathered from various in-house and online sources. These images were controlled for valence and familiarity. Of those images, 230 (115 faces, 115 scenes) were selected for use in this experiment. For each participant, 40 faces and 40 scenes were randomly selected to serve as the PM targets, 75 faces and 75 scenes were used as distractors. PM target images did not appear as distractors and were used on one trial only. Distractor images never reappeared within the same trial, but later reappeared on subsequent trials (mean exposures per distractor = 14, min = 6, max = 20).

PM Task Description

Each trial began with the presentation of the PM target (a face, a scene, or no target) for 3-sec, followed by a 1-sec fixation cross. For non-PM trials, participants saw a yellow null (∅) sign in lieu of a face or scene. For PM trials, participants were informed

that the PM target shown was only relevant for the current trial. After target presentation, participants saw a series of 1-15 probes per trial. Every probe contained a visual-search array in the center of the screen, with one face and one scene (each of size 9.5° x 9.5° in visual angle) vertically aligned with the center of the images placed 11.5° above or below the search array. Each probe was on the screen for 2 sec, during which the participants were allowed 1.9s to respond to the presence or absence of the horizontal arrow in the OG task, or to indicate whether the PM target had reappeared (by pressing “3” for behavioral participants or the third button on the button-box for MRI participants). Participants were instructed to equally weight the importance of both tasks, and only one response (to either the OG task or the PM task) was allowed per probe (“task-switch” approach, Bisiacchi, Schiff, Ciccola, & Kliegel, 2009). Visual feedback was presented immediately following each response, in the form of the arrows turning green for correct OG responses, turning red for incorrect OG responses, or a yellow border surrounding the screen for PM false alarms. Probe feedback remained on screen for the remaining duration of each 2-s probe. The 1.9s response deadline ensured that some time (minimum 100ms) was always devoted to feedback on every probe. On trials with a PM target presented at the beginning of the trial (“PM trials”), participants performed both tasks as described above. On trials with no PM target presented (“non-PM trials”), participants were instructed to ignore the face and scene images and focus solely on the OG task. The PM target reappeared only once per trial, and its reappearance always marked the end of the trial. After the final probe of each trial, participants were given feedback on the PM task in the form of a green border appearing around the edge of the screen for correct PM responses and a red border for

missed PM targets. This feedback (or a blank screen for non-PM trials) remained for 2s and was followed by a 6s ITI with a fixation cross on the screen before the next trial began.

PM and non-PM trials were randomly intermixed within each block, with one-third of all trials being non-PM trials, and the other two-thirds were equally split between face- or scene-target PM trials. Participants were able to rest between blocks for as long as they wanted before continuing with the experiment. OG task difficulty was manipulated in five conditions: it could either (1) increase starting at the easiest difficulty (level 1), (2) increase starting at the median difficulty (level 8), (3) decrease starting at the hardest difficulty (level 15), (4) decrease starting at the median difficulty, or (5) remain fixed at the median difficulty. For the main analyses reported here, the first two conditions were combined as "*increasing*" trials, the second two conditions were combined as *"decreasing"* trials, and the fifth condition was referred to as *"fixed".* Each of the five difficulty types occurred three times throughout each block in pseudo-random order. Fifteen "real" trials (non-"catch" trials) corresponding to the 3x5 combinations of PM type (face/scene/non-PM) and OG difficulty type occurred once per block. Five trials in each block were "catch trials" (containing fewer than 8 probes), which were included to keep participants engaged at the beginning of each trial. The difficulty types and target category for catch trials were counterbalanced across the entire experiment. Trial lengths were predetermined and pseudo-randomized so that every participant had the same number of total probes for face-target, scene-target, and non-PM trials. Scene and face locations were randomized on each probe, and faces and scenes appeared on the top or bottom of the display with equal probability. Performance on face and scene

trials was similar, so to increase statistical power, these trials were collapsed for subsequent analyses.

Changes in difficulty occurred at the rate of one shift in difficulty level per probe until either the end of the trial or until a difficulty endpoint was reached. For example, on a 10-probe trial starting at the easiest difficulty (level 1) and then increasing in difficulty, the trial would end at difficulty level 10. However, on a 10-probe trial starting at the median difficulty (level 8) and then increasing in difficulty, the highest difficulty (level 15) would be reached by the 8th probe, and the difficulty would remain at this level for the final two probes of the trial. For the behavioral experiment, this led to 70 PM probes at the hardest (15) and easiest (1) difficulty levels, 210 PM probes at the middle difficulty (8), and 30-40 PM probes at the other difficulty levels across the entire experiment. There were approximately half that many non-PM probes at each level. For the fMRI sample, in which there was one less experimental block, participants had an average of 53 probes at the hardest and easiest difficulties, 169 at the middle difficulty, and between 22-34 probes at the other difficulties. For non-PM probes, the count across the experiment was approximately half of those totals.

Ongoing task localizer (MRI sample only)

Once in the scanner, each participant first completed the OG arrow visual search task. The OG task localizer provided participants with the chance to familiarize themselves with this task in the MRI setting before the addition of the embedded PM task. It also allowed for us to collect timepoints where only the OG task was displayed on the screen for training the classifier on non-PM probes. The arrow search task was

the same as during the PM task, however there were no face or scene images on the screen during the OG task localizer. Participants indicated the presence or absence of a target right-facing horizontal arrow within a 1.9s response window. Feedback was given in the form of arrows turning green for correct responses and red for incorrect responses through the end of the response window. Similar to the main task, the difficulty of the OG task gradually increased or decreased over the course of a trial. Each block included 60 probes, split into eight trials of various lengths (min=2, max=12). There was a 6s ITI between trials. Participants performed two blocks of the ongoing task localizer (120 probes total) before moving on to the main PM task.

Face/scene localizer (MRI sample only)

After completion of the main PM task and a quick subsequent memory test (not discussed here), participants performed a face/scene sub-categorization localizer task. This localizer allowed for us to collect more data points of face and scene processing for training the fMRI pattern classifiers. Presentation of faces and scenes alternated in mini-blocks, where 11 stimuli from the same category (faces or scenes) were presented in a row, and participants indicated whether a face was male or female or whether a scene was indoors or outdoors. A single face or scene image was presented in the center of the screen, and participants had 1s to respond. Because of the short response window, female/outdoor was always presented as the left option, and male/indoor was always presented as the right option. Note that this corresponded to button box responses “1” and “2” in the scanner. Immediately following a response, a red (incorrect) or green (correct) box appeared around the selected choice, or a blue box (no response)

appeared around the correct choice for 500ms. The stimulus remained on the screen during this time. There were then 500ms between trials during which a fixation cross appeared on the screen. There were three blocks with six mini-blocks pseudo-randomly interleaved for a total of 198 localizer trials. While the fMRI pattern classifiers performed well without the inclusion of the data from both localizers, they performed numerically better when localizer data was included. Therefore, our final analysis included samples from both the OG localizer task as well as the face/scene localizer task.

MRI acquisition and preprocessing

MRI data were acquired on a 3.0-T Siemens Skyra MRI scanner with a 32-channel head coil. Whole brain, high-resolution anatomical images were collected for registration and parcellation using a T1-weighted MPRAGE sequence (repetition time (TR) = 1900ms, echo time (TE) = 2.43ms, flip angle = 9°, field of view (FOV) = 256x256x192, 1mm isotropic voxels). Functional images were acquired using a T2*-weighted multiband accelerated EPI pulse sequence (TR = 2s, TE = 29ms, flip angle = 78°, FOV=76x76, slice-thickness = 3mm, multiband factor = 2, number of slices = 48, no gap). Following shim adjustment at the beginning of the scan session, a B0 field-map with the same slice prescription as the functional data was acquired.

As an initial step for preprocessing, DICOMs were converted to NIFTI format using dcm2niix (Li, Morgan, Ashburner, Smith, & Rorden, 2016). Next, Freesurfer's (Fischl, 2012) recon-all function was used to skull strip and parcellate the high-resolution anatomical image. MRI data was preprocessed using a combination of Freesurfer 6.0, FSL 5.0.9 (Jenkinson, Beckmann, Behrens, Woolrich, & Smith, 2012),

and ANTs 2.1.0 (Avants et al., 2011) functions. Functional images were first slice time corrected using FSL's slicetimer function. Then, functional runs were normalized to the third run of the main task (middle run) using a combination of within-run motion correction, rigid and affine registration, and field unwarping processes from FSL and ANTs. Non-linear registration, via antsRegistrationSyn, was then used to correct for between run differences. Lastly, high pass filtering (128s) was applied to the images.

For the univariate GLM analysis, rigid and affine transformations were used to register the functional scans to the high-resolution anatomical, and then non-linear transformations were applied to normalize runs to the MNI-template. These images were spatially smoothed (5 mm gaussian), but no further preprocessing was performed before using FSL FEAT for modeling. The model included separate regressors for the PM target presentation, PM trial probes 1 to n-1, probe n (when the PM target reappeared in PM trials), non-PM trial probes 1 to n, trial-feedback, and rest. Six motion parameters, extracted using FSL's MCFLIRT, were included as confound regressors. FSL's FEAT was used to identify voxels that were more responsive on PM trials (PM+OG) than on non-PM (OG-only) trials (cluster correction, $p < .001$; **Figure 3A**; for ROI list see **extended data Figure 3-1**).

This group level results map was then individually transformed from standard space into subject functional space so that the multivariate pattern analysis could be performed independently for each participant. To create subject-specific masks we performed B0-field unwarping, rigid, and affine registration, and lastly non-linear registration of EPIs to a middle functional run using FSL and ANTs registration tools, and then co-registered EPI volumes to their own MPRAGE structural volume using a

similar process of linear transformations. We then used a combination of FSL FMRIB's Linear Image Registration Tool (FLIRT) and ANTs non-linear registration tool antsRegistrationSyN to register structural volumes to MNI space. Individual, native-space masks were created by combining (with the registration parameters for the MPRAGE) and applying a reversed transformation matrix from EPI to MNI stereotaxic space on the group-level GLM mask described above.

The classification analysis was performed using the Princeton MVPA toolbox in MATLAB, https://github.com/princetonuniversity/princeton-mvpa-toolbox. Binary (one vs. all other classes) L2-penalized logistic regression classifiers (penalty=50) were trained, separately for each participant, to differentiate fMRI activity corresponding to faces, scenes, the OG task, and resting periods in between trials. A combination of data from the OG task localizer (labeled as class "OG"), from the face/scene localizer (labeled as class "face" or "scene"), and from the PM task (labeled as "face", "scene", or "OG" depending on the PM target for each trial) were used to train the classifiers using k-fold cross validation. From the main PM task, data from all probes (except for the final probe on each trial when the PM target reappeared) were used for training. For the k-fold cross-validation procedure, classifiers were trained on all data from the two localizer tasks plus 4 of the 5 runs from the PM task. In total, there were 7180 timepoints used for classifier training on each iteration (face = 1695, scene = 1695, OG = 1900, rest = 1890), and 1215 used for testing on each iteration (face = 300, scene = 300, OG = 325, rest = 290). All regressors were shifted forward in time by two TRs (4 s) to account for the hemodynamic lag. These classifiers were then applied to the held out run of data from the PM task. The PM task runs were then rotated, and this procedure was

repeated to train classifiers and then test them on the next held-out task run. This procedure was performed five times so that all runs of the PM task were tested. To improve classifier accuracy, we performed feature selection to remove uninformative voxels from the training data. This was done separately for each fold of the cross-validation analysis. Data within each voxel were z-scored across all timepoints, and a 1x4 ANOVA was performed to select only those voxels that demonstrated significant ($p<.05$) variance across the four classes being trained (face, scene, OG, and rest). The mean number of voxels selected for each participant was 7864 voxels (SEM = 1208).

Each of the 1-vs-other classifiers produced a class evidence score for the class on which it was trained. Therefore, the four evidence values produced for each test timepoint need not sum to one. At each timepoint, the class with the highest evidence value was selected as the predicted output. These predicted outputs were compared to the actual class of the timepoint (face, scene, OG, or rest) in order to calculate the classifier accuracy. AUC was calculated in MATLAB by comparing correct predictions and false alarms independently for each category across all time points. Scrambled regressor assignments were used to test the empirical chance level of classifier performance trained in this way (n = 1000 per participant). Average scrambled baseline performance (mean = 27.52%) was similar to the empirical chance level of 25% for all participants.

Calculating PM cost and PM cost slope

For all analyses involving RTs, we excluded any responses faster than 300ms. This criterion is in line with previous work (Boywitt & Rummel, 2012; Horn, Bayen, &

Smith, 2013) and was used in order to exclude late responses carried over from the preceding response window. We hypothesized that individuals would be able to adjust their PM strategy on a moment-to-moment basis in response to fluctuating cognitive demands. To initially test this theory, we performed a second pilot study where we had participants perform our PM task at either an easy (level 4 in the main study) or hard (level 12 in the main study) OG difficulty level. We found that PM cost significantly varied as a function of OG task difficulty ($F(1,19) = 35.63$, $p<.001$), with cost being higher at the easier difficulty (easy PM cost mean = 0.134 s (SE = 0.012)) than at the harder difficulty (hard PM cost mean = 0.031 s (SE = 0.012)). PM Accuracy was equivalent across difficulties ($F(1,19) = 0.785$, $p = 0.387$; easy = 71.0% (4.5%), hard = 64.5% (5.8%)). For more information about pilot study two, see **extended data Figure 2-1**. In the current study, we calculated the "PM cost" at each task difficulty level associated with making a correct response on the OG task with vs. without the additional demand of the PM task (i.e., PM trials vs. non-PM trials).

In order to calculate a PM cost for each probe, we first calculated the average OG RT on non-PM probes at each difficulty. We performed this analysis separately for each participant. We included only correct OG responses, however a control analysis including both correct and incorrect OG task responses produced qualitatively similar results. There was a practice effect of decreasing overall RTs between early experimental blocks and late experimental blocks ($F(1,77) = 87.1$, $p < .001$). To account for these practice effects, we calculated the non-PM OG RT baseline separately for the first half (early) and second half (late) of the experiment. For the behavioral sample in experiment 1, early trials came from blocks 1-3 and late trials from blocks 4-6. The MRI

sample in experiment 2 had one practice block outside of the scanner (data not recorded), so for those participants early trials came from main task blocks 1-2 and late trials from blocks 3-5. This protocol created 30 baseline non-PM OG RT values for each participant: 15 difficulties x 2 time bins (early/late). To create the PM cost for each individual probe from PM trials, we first determined the relevant baseline value (the value with the same difficulty and time bin as the probe) and subtracted that value from the OG RT on that probe. This enabled us to estimate the PM cost for every probe of the experiment. After obtaining PM cost values for each probe, we then calculated the degree to which PM cost shifted over the course of each trial (*PM cost slope*). PM cost slope was determined by calculating the difference in average PM costs of the first three probes vs. the final three probes of a trial (excluding the very last probe in which the PM target appeared), and then dividing by the number of probes in the trial. In order to account for lapses in motivation and to exclude trials where there was a lack of correct OG probes for calculating a PM cost slope, we excluded any trial where OG task accuracy fell below the chance level of 50% (this led to an average of 9.9% (SE = 1.0%) of trials per participant being excluded).

Because this paradigm involves long PM trials (mean 28s, range 8-36s), it produced few behavioral PM reports per participant. Therefore, to increase statistical power for the analysis used to evaluate the relationship PM accuracy and PM cost slope and PM intention evidence, we performed a non-parametric bootstrap analysis (Efron, 1979) using data sampled from all participants (see Kim, Daselaar, & Cabeza, 2014; Lewis-Peacock, Cohen, & Norman, 2016 for previous use of this method). On each bootstrap iteration (n=10000) of this analysis, 78 participants were selected at random

with replacement and combined to make one super subject. By using random selection with replacement, this form of analysis allows us to see to what extent the results are generalizable (Thompson, 1993). We used logistic regression to test the relationship between PM accuracy and PM cost slope and/or PM intention evidence on the different trial types. The stability of the effects across all iterations was analyzed to assess population-level reliability.

Code Accessibility

The code/software described in the paper will be made freely available online at [URL redacted for double-blind review]. All code for this experiment was run using Psychophysics Toolbox Version 3 in Matlab 2014b on a 21.5" iMac computer with operating system OS X 10.11. All statistical analyses of behavioral data were performed using R (v 3.4.1, R Core Team, 2017). Programs used in neuroimaging analyses are listed in the methods section.

Results

Ongoing Task Performance

Participants performed well on the ongoing (OG) task (mean accuracy = 83.64%, 95% CI = [75.49%,91.79%]) as summarized in **Figure 2** (for a comparison between experiment 1 and 2, see **extended data Figure 2-2**). There was no interaction between trial type (PM trials versus non-PM trials) and task difficulty ($\beta_{interaction} = 6.2*10^{-4}$, 95% CI = [-0.001, 0.002], $p_{interaction} = 0.468$) on OG task accuracy. OG accuracy decreased as difficulty increased, and there was a small, but reliable main effect of trial type between

PM and non-PM trials (β_{diff} = -.026, 95% CI = [-0.027, -0.025], p_{diff} <.001; β_{pm} = 0.008, 95% CI = [0.001, 0.016], p_{pm} = 0.024; marginal r^2 = 0.57; **Figure 2A**). The main effect of trial type indicated a dip in OG accuracy of less than 1% on PM trials compared to non-PM trial. A follow-up analysis compared OG accuracies from PM and non-PM trials at each difficulty level, finding that accuracies were only reliably different at difficulty level 8 on fixed difficulty trials (p = 0.025, after Bonferroni correction factor 15). On dynamic trials (increasing and decreasing difficulty), the main effect of PM task on OG accuracy was not significant (β_{pm} = 0.007, 95% CI = [$-8.5*10^{-5}$, 0.015], p_{pm} = 0.052). We also found that at the hardest difficulty level of the OG task (15), participants were still performing well above chance (mean = 61.43, t(77) = 80.39, p<.001, 95% CI = [59.90%, 62.95%]).

One of the primary methods for inferring PM strategy use is measuring the difference in RTs between PM and non-PM trials on an ongoing task (Einstein & McDaniel, 2005). A large difference in OG RTs between PM and non-PM trials implies the use of proactive control on the PM trials, whereas a small difference implies the use of reactive control on the PM trials. Here, we found an interaction in OG RTs between condition (PM and non-PM trials) and difficulty level (1 to 15) ($\beta_{interaction}$ = 0.011, p < .001, 95% CI = [0.010, 0.013]; **Figure 2B**). At the hardest difficulty level, RTs were well below the response deadline of 1900ms (t(77) = 38.401, p<.001, mean RT = 1264ms, 95% CI = [977ms, 1551ms]), demonstrating that participants were performing above floor. The difference in OG RT between PM and non-PM trials (referred to as “PM cost” from here on out) was then calculated for each participant at each difficulty level. These data were replotted in this fashion, and they reveal that average PM costs varied systematically as

a function of OG task difficulty (β_{cost} = -10.35, p < .001, 95% CI = [-12.29, -8.42]; **Figure 2C**). This suggests that PM strategy shifted flexibly between proactive control and reactive control as the OG task increased in difficulty, and vice versa.

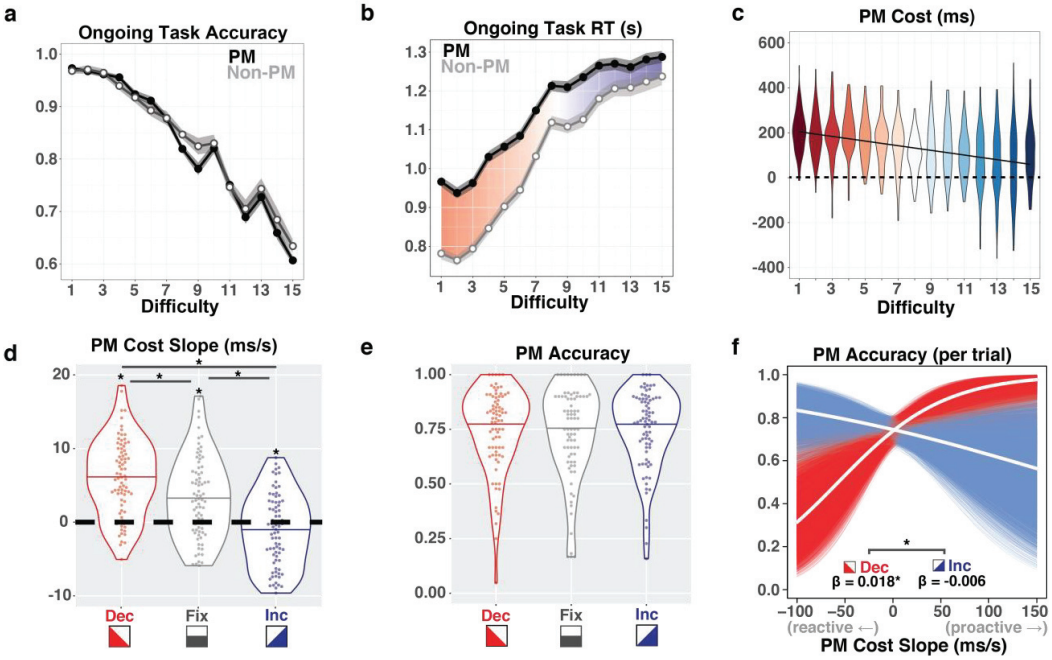

Figure 2. Behavioral performance. a) Ongoing task accuracy across difficulties, error ribbons +/- 1 SEM. PM: dual-task trials with a PM intention; Non-PM: ongoing task only without a PM intention. b) Ongoing task RT (correct responses only) across difficulties, error ribbons +/- 1 SEM. c) PM cost (the difference between ongoing task RT for PM trials vs. Non-PM trials) was computed for each participant at every difficulty level. Violin plots represent the distribution of by-participant average costs at each difficulty. PM cost is highest at easy difficulty levels (dark red) and decreases as task difficulty increases (dark blue). d) Polynomial model fits validated the use of linear models which allowed us to calculate the shift in PM cost on each trial (see text below and **extended data Figures 2-1** and **2-2** for further analysis details). Next, PM cost slopes were calculated as the change in PM cost within each trial. Violin plots show the average within-trial PM cost slopes for decreasing (Dec), fixed (Fix), and increasing (Inc) trials across participants. * p < .05. e) PM accuracy for each participant across trial types. f) Logistic regression bootstrap analysis linking PM cost slope to PM accuracy for decreasing (red) and increasing (blue) trials. Each individual red/blue line shows the predicted relationship for each bootstrapped sample (n=10,000). White lines reflect the fixed-effects relationship for the original sample. * p < .05.

To evaluate whether this linear shift in PM strategy across task difficulty levels held within individual trials, we computed 1st, 2nd, and 3rd order polynomial regressions between PM cost and OG difficulty separately for each trial. If PM strategy selection was bimodal (i.e., an all-or-none "flip" between proactive and reactive control), then individual trials should be best fit by a 3rd order polynomial. However, if the relationship was more fluid, then a 1st or 2nd order polynomial should fit the data better. Additionally, if the data demonstrate a dramatic U-shaped or asymptotic curve, instead of a linear fit, a 2nd order polynomial should fit better than a 1st order polynomial. We compared Akaike information criterion values (AICc, with a correction for small sample sizes) for each model for correct responses on each trial and used Akaike weighting to compare relative model fits. We found that a 1st order polynomial (linear model) fit best for nearly all trials (see **extended data Figure 2-3** for data visualization: mean = 93.43%, 95% CI = [92.62%, 94.24%]; Akaike weight = 0.873, 95% CI = [0.865, 0.880]), compared to 2nd order fits (mean = 5.72%, 95% CI = [4.99%, 6.45%]; Akaike weight = 0.111, 95% CI = [0.104, 0.119]), or 3rd order fits (mean = 0.85%, 95% CI = [0.68%, 1.02%]; Akaike weight = 0.016, 95% CI = [0.012, 0.019]). One concern with fitting models on a by-trial basis is that noise may bias model selection towards the models with fewer parameters. In order to address this concern, we performed a less conservative AIC (without the small sample correction term) model selection analysis and then a separate bootstrap analysis where polynomial fits were calculated for a random sub-sample of trials. Both of these analyses corroborate our original finding and indicate that a linear fit is the most likely descriptor for a majority of trials (see **extended data Figure 2-1**). This result

provides evidence that the engagement of different control strategies often changes fluidly and linearly in accordance with shifts in OG task demands within a PM trial.

Next, we evaluated whether changes in PM cost within a trial were different for increasing-, decreasing-, and fixed-difficulty trials. To do this, we computed a within-trial measure of linear shift in PM cost from the beginning of the trial to the end of the trial, which we shall refer to as *PM cost slope*. We found that PM cost slopes varied systematically across trial types ($F(2,154) = 47.02$, $p < .001$; **Figure 2D**). PM cost slopes were negative on increasing trials (mean = -1.14 ms/s, 95% CI = [-2.24,-0.42], $t(77) = 2.07$, $p = 0.042$), and were positive on decreasing trials (mean = 6.03ms/s, 95% CI = [5.44, 6.62], $t(77) = 10.23$, $p<.001$) and fixed trials (mean = 3.34ms/s, 95% CI = [2.07, 4.61], $t(77) = 5.24$, $p< .001$). In this task design, a PM target appeared at the end of every PM trial. Thus, positive PM cost slopes (i.e., a shift towards proactive control) on fixed difficulty trials likely arose from an increase in PM expectancy as each trial progressed (Bowden, Smith, & Loft, 2017; Oksanen, Waldum, McDaniel, & Braver, 2014). Although expectancy should impact all trial types equally, there were meaningful differences between conditions. Planned pairwise comparisons revealed that PM cost slopes increased step-wise from increasing trials to fixed trials ($F(77) = 37.14$, $p < .001$) and from fixed trials to decreasing trials ($F(77) = 15.1$, $p<.001$).

Prospective Memory Task Performance

On average, participants identified the PM target on three-quarters of the trials (mean PM accuracy = 74.57%, 95% CI = [56.82%, 92.31%], with no differences in accuracy across trial types ($F(2,154) = 0.679$, $p = 0.508$; **Figure 1E**). The false alarm

rate, defined as PM-target responses on probes where the target was not present, was low (mean = 0.60% of probes, 95% CI = [-0.72%, 1.92%]). Because PM accuracy was stable across the dynamic trial types (increasing and decreasing difficulty trials), and OG accuracy was not impacted by the presence of the PM task on these trials, we concluded that participants were not sacrificing accuracy on one task in order to perform the other. Therefore, any RT differences on the OG task during PM trials could reasonably be interpreted as reflecting differences in strategy used to perform the PM task, rather than a speed/accuracy tradeoff between the dual tasks.

Linking Shifts in PM Strategy to PM Performance

We hypothesized that not only would individuals demonstrate gradual shifts in PM strategy in response to changing task demands, but that changes in PM strategy would be related to PM performance. Based off of the dual methods of control framework (DMC, Braver, 2012) and dynamic multiprocess view of prospective memory (DMPV, Scullin et al., 2013), we reasoned that on decreasing and fixed difficulty trials, when the resources to implement proactive control became readily available, proactive control would benefit PM performance. However, on increasing difficulty trials, ongoing task demands make it difficult to implement the proactive control mechanisms of strategic monitoring and/or sustained representation of the PM-target. Therefore, individuals may benefit from shifts towards reactive control strategies on these trials, as they attempt to preserve performance on the PM task in the face of increasing demands. To evaluate this hypothesis, we tested the relationship between *PM cost slope* and *PM accuracy* across all trial types.

We used bootstrapped logistic regression to relate these two measures separately for increasing and decreasing trials (**Figure 2F**). On decreasing trials, larger *positive* PM cost slopes (reflecting a shift towards proactive control) were related to better PM performance (β_{dec}= 0.018, p < .001, 95% CI = [0.008, 0.028]). On increasing trials, larger *negative* PM cost slopes (reflecting a shift towards reactive control), were numerically related to better PM performance, but this relationship did not reach statistical significance (β_{inc}= -0.006, p = 0.119, 95% CI = [-0.015, 0.004]). Critically, there was an interaction of PM cost slope and trial type on PM accuracy, with the direction of PM cost slope leading to different consequences on increasing vs. decreasing trials ($\beta_{interaction}$= 0.024, p < .001, 95% CI = [0.011, 0.037]). On fixed trials (data not shown), the relationship between PM cost slope and PM accuracy was positive (β_{fix}= 0.020, p = 0.003, 95% CI = [0.006, 0.033]), which was similar to decreasing trials (β_{diff} = 0.001, p = 0.423, 95% CI = [-0.015, 0.017]) but significantly more positive than on increasing trials (β_{diff} = 0.025, p = 0.001, 95% CI = [0.009, 0.041]).

The relationship between PM cost slope and PM accuracy survived after controlling for other possible explanatory variables. We first compared the by-trial Akaike information criterion (AIC) and variance explained (R^2) values for predicting PM accuracy on increasing and decreasing trials using five different factors: average OG RT, average OG accuracy, OG RT slope, average PM cost, and PM cost slope. We ran a bootstrap analysis (n=3,000 iterations) comparing models that included the interaction of trial direction and each of these factors for predicting PM accuracy, and we extracted AIC and R^2 scores from each iteration. The AIC scores were converted to Akaike weights (Wagenmakers & Farrell, 2004) for comparison across models. The model

using PM cost slope as a predictor of PM accuracy had a significantly higher Akaike weight (mean = .99, SE = .001) than any other model (all other means < .01). The R^2 value for the model containing only PM cost slope was also the highest (mean R^2 value: PM cost slope = 0.08, average PM cost = 0.03, average OG RT = 0.03, average OG Acc = 0.005, OG RT slope = 0.005).

To see if this relationship held at the subject level, we also performed partial regressions comparing how much PM cost slope predicted PM accuracy while controlling for the other variables. After controlling for average OG RT, average OG accuracy, average PM cost, and OG RT slope one at a time, PM cost slope still explained a significant proportion of variation in PM accuracy (R^2 = .09, p=0.009; R^2 = 0.12, p = 0.002; R^2=.12, p = .001; R^2=.12, p = .002; respectively).

This same by-trial relationship between PM cost slope and PM accuracy existed across participants as well. Participants who on average showed larger shifts towards proactive control (more positive PM cost slopes) benefited more on decreasing trials, and participants who showed larger shifts towards reactive control (more negative PM cost slopes) benefited more on increasing trials ($\beta_{interaction}$ = -0.018, p = 0.002, 95% CI = [-.013, -.024], data not shown). In summary, these results provide behavioral evidence that individuals shifted PM strategy from moment to moment in response to changing OG task demands. These shifts in cognitive control were adaptive because their direction and magnitude were related to successful PM performance.

Neural Measures of PM Intentions

For the participants who performed this task in the MRI scanner (N=28), we evaluated whether a neural measure of PM intention-related brain activity (Lewis-Peacock et al., 2016) could provide additional insight into the link between PM strategy selection and memory performance. First, we identified regions that were significantly engaged by the PM task above and beyond the OG task in isolation (**Figure 3A**). These regions are consistent with previous literature on PM intention maintenance (Beck et al., 2014; Cona et al., 2015; Lewis-Peacock et al., 2016; McDaniel et al., 2013). From these brain regions, fMRI pattern classifiers were used to quantify the degree of PM intention-related processing across each trial. The strength of PM intention processing was operationalized as the difference in classifier evidence for the PM-relevant category vs. the PM-irrelevant category (e.g., "face *minus* scene" for a face-target PM trial). Trained classifiers performed well above chance at predicting the category of the current trial's PM target (classifier AUC for faces: 86.94, SE = 0.01; scenes = 88.54, SE = 0.01; non-PM trials: 83.76, SE = 0.01; and rest = 99.62, SE = $4.7*10^{-4}$). Across trials, this neural measure varied systematically with OG task difficulty (**Figure 3B**). The neural evidence of PM intention processing decreased as task demands increased (β = -0.005, p <.001, 95% CI = [-.008, -.002], marginal r^2 = .039).

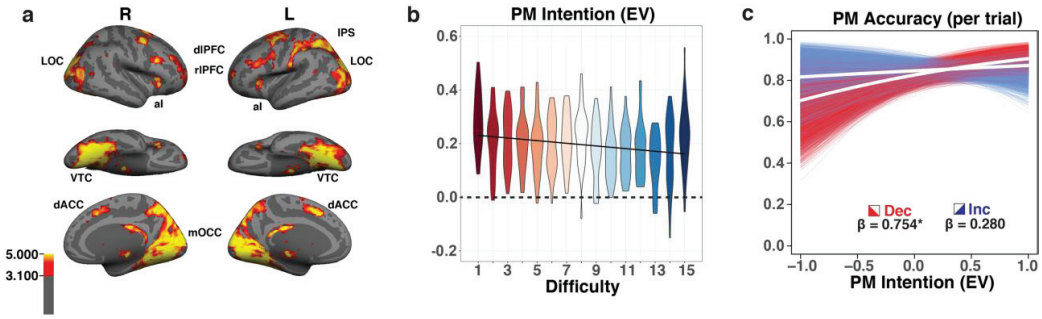

Figure 3. fMRI decoding of PM intentions. a) Brain regions significantly engaged by the addition of the PM task to the OG task (GLM contrast PM > Non-PM, FDR corrected at p<.001; for ROI list see **extended data Figure 3-1**). aI = anterior insular cortex, dACC = dorsal anterior cingulate cortex, dlPFC = dorsal lateral prefrontal cortex, IPS = Intraparietal sulcus, LOC = lateral occipital cortex, mOCC = medial occipital cortex, rlPFC = rostral lateral prefrontal cortex, VTC = ventral temporal cortex. These regions were used as the initial feature mask to train and test fMRI pattern classifiers for PM intention-related activity. To more directly identify regions primarily responsible for PM-intention representation during this task, we performed a surface-based searchlight analysis. That analysis indicated that the VTC and LOC were more important for PM-processing. (for more details see **extended data Figure 3-2**) b) PM intention evidence (EV; the difference between classifier evidence for the PM target's category and the non-target category) was computed for each participant at every difficulty level, and group data is shown in violin plots. PM intention evidence was highest at easy difficulties (dark red) and lowest for the most difficult levels (dark blue). c) The relationship between PM intention evidence and PM accuracy was computed using bootstrapped logistic regression (n=10,000 iterations) for decreasing (red) and increasing (blue) trials. * p < .05.

However, within trials the neural measure of PM intention processing did not vary systematically across timepoints (mean slope of PM intention evidence = 0.007 EV/s, 95% CI = [-0.002, 0.016], t(27) = 1.57, p = 0.128). There were also no differences in by-trial PM intention evidence slopes across increasing, decreasing, and fixed difficulty trials (F(2,54) = 1.35, p = 0.269). The stable level of PM intention processing over the course of a single trial may be a measurement limitation due to the temporal sluggishness of the BOLD signal. Alternatively, it could also reflect the engagement of a prospective "retrieval mode" (Cona, Bisiacchi, & Moscovitch, 2014; Guynn, 2003; 2008),

which has been described as a more sustained and relatively inflexible component of proactive control that involves PM-items being held in some prioritized state of working memory (Underwood et al., 2015). Therefore, we computed the *average* classifier evidence for the PM intention on each trial and related this (rather than the slope) to PM accuracy. A mixed-effect ANOVA confirmed that there were no overall differences in average PM intention evidence across trial types ($F(2, 54) = 0.40$, $p = 0.670$). This result was expected because increasing and decreasing trials spanned the same range of difficulty levels (e.g., 1 to 15 vs. 15 to 1). We found that average PM intention evidence correlated positively with PM accuracy on decreasing trials ($\beta_{dec} = 0.754$, $p = .017$, 95% CI = [0.047, 1.50]; **Figure 3C red**) and fixed trials ($\beta_{fix} = 0.976$, $p = .039$, 95% CI = [-0.121, 2.062]; not shown), but not on increasing trials ($\beta_{inc} = 0.280$, $p = 0.255$, 95% CI = [-0.679, 1.123]; **Figure 3C blue**). There were no reliable differences in this statistic between increasing trials and either decreasing trials ($\beta_{interaction} = 0.474$, $p = 0.199$, 95% CI = [-0.627, 1.623]) or fixed trials ($\beta_{interaction} = 0.223$, $p = 0.361$, 95% CI = [-1.119, 1.473]).

Combining Behavioral & Neural Measures to Predict PM Performance

We sought to test whether combining both the time-sensitive but indirect behavioral metric of PM cost slope (putatively reflecting dynamic shifts in PM strategy) and the coarser but more direct neural metric of PM intention evidence (putatively reflecting sustained PM engagement) could improve our prediction of PM accuracy on a trial by trial basis. There was no by-trial correlation between these measures (mean $r = 0.02$, 95% CI = [-0.36, 0.39], $p = 0.92$), suggesting that the two metrics could provide

unique information about task performance. We performed a bootstrap analysis to calculate the Akaike information criterion (AIC) values for models predicting PM accuracy including all possible combinations of the predictors *PM cost slope*, *PM intention evidence*, and *trial type* (increasing/decreasing). We then selected the best performing model that included (1) a neural and a behavioral metric, (2) only a behavioral metric, or (3) only a neural metric. Next, we converted AIC scores for these three models to Akaike weights (wAIC), allowing us to directly compare AIC values as conditional probabilities (Wagenmakers & Farrell, 2004). The results show that the combined Behavior & Neural model (the full model including all main effects, all two-way interactions, and the three-way interaction of PM cost slope, PM intention state, and trial direction) was the best model (Wilcoxon median Akaike weight = 0.889, Wilcoxon 95% CI =[0.883,0.896], p< .001; **Figure 4A**). This combined model was significantly more likely than either the best Behavior-only model (Wilcoxon median ratio = 149.91, wilcox 95% CI = [135.47, 165.63], p < .001) or the best Neural-only model (Wilcoxon median ratio = $4.5*10^9$, Wilcoxon 95% CI = [$3.9*10^9$, $5.2*10^9$], p < .001).

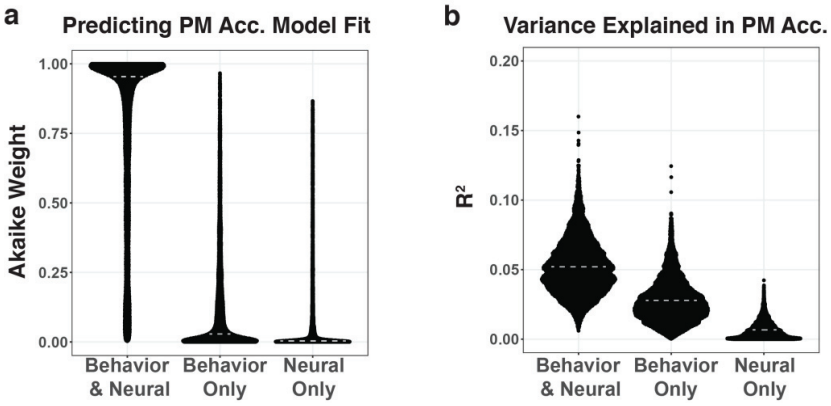

Figure 4. Model comparisons using behavioral and neural metrics to predict PM performance. Model weights and R^2 values were computed across bootstrap iterations

(n = 10,000) to test model differences. a) Akaike weights (wAIC) across bootstrap iterations for each model. b) Explanatory power of each model (R^2) shown as distributions across bootstraps. Medians are indicated by dashed grey lines.

The Behavior & Neural model explained the most variance in PM accuracy, with an R^2 value of approximately double that of the best Behavior-only model (R^2 = .052 vs. .028, **Figure 4B**). Having found that the full Behavioral & Neural model was the best predictor of PM performance, we further investigated the reliability of each relationship in the combined model using the bootstrap approach. In this analysis, we were also interested in whether the relationships we found independently between PM accuracy and PM cost slope and then between PM accuracy and PM intention evidence would be qualified by any reliable interactions in the full model. We found that although including the three-way interaction term and multiple two-way interaction terms resulted in the lowest overall AIC score, the only statistically reliable interaction was that of *PM cost slope* and *trial direction* (p = 0.026), confirming that the analysis shown in **Figure 2F** holds in a more comprehensive model. Additionally, we found that PM intention evidence had a reliably positive relationship to PM accuracy on decreasing difficulty trials (mean = 0.027, p = 0.024), but that relationship was not reliable on increasing difficulty trials (mean = -0.003, p = .395), also confirming the analysis shown in **Figure 3C** holds in a more comprehensive model. In summary, our model tests revealed three main results: 1) Including both the neural and behavioral metrics of proactive control improved prediction of PM accuracy over using either metric independently. 2) PM cost slope was differentially predictive of PM accuracy for increasing vs. decreasing trials, replicating the relationship from our behavior-only analysis above. 3) Higher levels of

PM intention evidence were positively related to PM performance on decreasing difficulty trials, but there was no reliable relationship on increasing difficulty trials.

Discussion

This study investigated how navigating an environment with rapidly shifting cognitive demands impacts how we remember to perform future actions. A delayed-recognition prospective memory (PM) task was combined with a dynamic visual search ongoing (OG) task that varied in difficulty from moment to moment. When task difficulty was low, there was greater behavioral interference from the PM task ("PM cost": slower RTs in the OG task) and stronger neural representation of the PM intention ("PM intention evidence": classifier evidence for the PM target category in PM-sensitive brain regions). Both of these measures reflect components of proactive control (Braver, 2012), and were negatively correlated with OG task difficulty. The behavioral measure varied within a trial according to the task demands, whereas the neural measurement was stable within a given trial but varied across trials. Combining these behavioral and neural measures provided the best prediction of PM accuracy from trial to trial. Together, these results suggest that individuals dynamically adjust their PM strategy in response to changes in environmental demands. Critically, we found that these shifts in PM strategy were adaptive because greater shifts (in the appropriate direction towards proactive or reactive control) were related to improvements in PM performance. The present results demonstrate that the ability to flexibly adjust cognitive control strategies, in response to changes in environmental demands, is an important contributor to successful execution of delayed intentions.

We computed two distinct metrics of proactive strategy use: a time-sensitive behavioral measure of PM cost, and a more tonic neural measure of PM intention processing. The amount of PM costs (the behavioral measure) has been repeatedly linked to levels of strategic monitoring for the PM intention (Einstein & McDaniel, 2005; Smith, 2003). Here, we found that *changes* in the amount of PM costs over the course of a trial were associated with better performance (**Figure 2F**). The dual mechanisms of control (DMC) framework suggests that proactive control would be favored on decreasing difficulty trials, when the OG task becomes progressively easier, because attention and working memory resources should be readily available to accomplish both the OG task and PM task successfully. On these trials, strategically monitoring for the PM intention may be worth the extra cost incurred in RTs on the OG task. Our results support this idea, showing that when participants reallocated cognitive resources to use proactive control on the PM task (positive PM cost slopes within a trial), PM performance improved.

However, we found that in increasing difficulty trials, there was an opposite relationship between PM cost slope and PM accuracy, where larger PM cost slopes were related to moderately worse PM performance. The DMC framework predicts that as difficulty increases, the ability to strategically monitor for the PM intention can be compromised, and reliance on proactive control may lead to deficits in PM performance. Such deficits may arise from interference in working memory caused by failed attempts to maintain a robust representation of the PM target in the face of distractors, a reduced ability to shift attention between the two tasks in order to strategically monitor for the PM cue effectively, or a reduced ability to perform the PM intention even after noticing a

prospective cue (Ballhausen et al., 2017; West, Carlson, & Cohen, 2007; Zuber, Kliegel, & Ihle, 2016). Consistent with these ideas, we found that when participants attempted to maintain high levels of proactive control even as the OG task difficulty increased (i.e. PM cost slopes were *positive* on these trials), PM performance suffered. The relationship between PM cost slope and PM accuracy on increasing difficulty trials suggests that reactive control can be used successfully in situations that are not well suited for proactive control (e.g., under high cognitive load). The results from this study build upon previous research that demonstrated in some circumstances there is a benefit to using proactive control (Shelton et al., 2016; Smith 2003), as we found on decreasing and fixed difficulty trials, and some circumstances where there is no benefit (for example: Einstein, 2005; Scullin et al., 2010), as we found on increasing difficulty trials.

In our study, participants knew that a PM target would reappear relatively soon after it was introduced (between 2-30s later with 100% fidelity). The DMPV framework (Scullin et al., 2013; Shelton & Scullin, 2017) posits that in contexts similar to our experiment, where PM occurrences are highly probable, individuals are biased towards and benefit from using proactive control. On trials with fixed difficulty, we found a consistent increase in PM costs across each trial (positive PM cost slopes), and greater increases in cost were related to better PM performance. This indicates a beneficial, perhaps default, shift towards proactive control in this paradigm for which there is an increasing probability of a PM event throughout each trial. Shifts towards proactive control were even stronger (and beneficial for performance) on decreasing difficulty trials as more cognitive resources became available over time. This result is in line with

previous work showing that given the available resources, individuals will increase monitoring as the expectancy of the PM event increases, and that increased monitoring is beneficial to PM performance (Bowden et al., 2017; Kuhlman et al., 2014; Loft, Bowden, Ball, & Brewer, 2014).

Additionally, even though overall PM accuracy on increasing and decreasing trials was equivalent, the range of accuracies differed between trial types. On decreasing difficulty trials, participants performed the PM task dramatically better when shifting towards proactive control and worse when shifting towards reactive control. On increasing difficulty trials, while this relationship was numerically reversed, the difference in performance across strategy types was reduced (see **Figure 2F**). In other words, on decreasing difficulty trials there was a clear and large benefit to PM performance when PM cost slopes were positive, while on increasing difficulty trials PM performance was more similar in respect to PM strategy. Additionally, when collapsing across all trials, we found a small performance advantage to using proactive control. On trials when the PM target appeared while PM costs were high, detection accuracy of the PM target was high (mean = 76.62%, SEM = 1.94%; average N = 37.7 trials/participant). On trials where PM costs were absent (indicating no evidence of proactive control) when the PM target appeared, accuracy was worse ($t(76) = 4.684$, $p < .001$), but still relatively good and well above floor (mean = 67.81%, SEM = 2.67%, N = 12.1 trials/participant). The behavioral data from our study suggests that over short, highly predictable intervals (< 30s), proactive control is the more reliable strategy for PM intention fulfillment, but only when monitoring for and maintenance of a PM intention can be adequately performed. In other circumstances, such as those with high

concurrent demands, individuals benefit from offloading the PM task to reactive control in the form of equal performance with less cost. Future work should investigate the impact of strategy flexibility across longer delays between encoding and retrieval of intentions, and whether a proactive benefit may still be observed when concurrent demands are high.

Many models of PM have focused on the relationship between proactive and reactive control. Some propose a central executive process that allocates resources either towards proactive or reactive control along a continuum (Gilbert, Hadjipavlou, & Raoelison, 2013), or strikes some balance between attention to external stimuli vs. internal stimuli (Cona, Scarpazza, Sartori, Moscovitch, & Bisiacchi, 2015). Guynn's (2003; 2008) two-component model of proactive control dissociates a flexible, strategic monitoring component from a more tonic component referred to as the "PM-retrieval mode", which involves sustained maintenance of the PM task set. This second component is described as load-invariant and relative inflexible, while the first component is thought to be dynamically sensitive to environmental factors (Underwood, Guynn, & Cohen, 2015). Recent work has suggested that the PM-retrieval mode is also able to be strategically adjusted to some extent (Whitehead & Egner, 2018), but the amount of cognitive resources needed to maintain it can negatively affect monitoring ability (Ballhausen, Schnitzspahn, Horn, & Kliegel, 2017).

In the present study, we propose that the behavioral measure of PM cost reflects the strategic monitoring component of this model, whereas the neural measure of PM intention evidence reflects the PM-retrieval mode component. Our results implicated neural regions commonly associated with working memory (Eriksson, Vogel, Lansner,

Bergström, & Nyberg, 2015, see **extended data Figure 3-1** for specifics) in supporting PM performance. We found that on decreasing difficulty trials (which, according to the behavioral analysis, favor a shift towards proactive control), PM intention maintenance was positively correlated with memory performance, but on increasing trials (which favor a shift towards reactive control) it was not (**Figure 3C**). Although individuals sometimes exhibited a high level of PM readiness, this did not influence task performance in situations that favored reactive control. These results are again in line with previous research suggesting that PM intention maintenance could be beneficial to PM performance in some situations but not *necessary* for successful realization of PM intentions in all situations (Cona et al., 2014).

We find that our behavioral and neural metrics provide complementary but independent information about PM performance, approximately doubling our model's predictive power when the neural measure of PM intention maintenance was combined with the behavioral metric of PM cost slope (**Figure 4B**). This result suggests that our measures are capturing different components of proactive control, though the current design does not specify the relative contribution of strategic monitoring versus PM intention maintenance. A future direction will be to use a more time-sensitive neural measure like EEG (electroencephalogram) and eye-tracking to measure active maintenance of PM intentions in dynamic environments, as well as to better identify late-retrieval mechanisms that are characteristic of reactive control strategy use.

Contrary to previous studies that have indicated a key role for the anterior prefrontal cortex in representing prospective intentions (Gilbert, 2011; Momennejad & Haynes, 2012, 2013), we found that regions known to support perception and working

memory for the PM intentions used here (i.e., the ventral temporal cortex (D'Esposito & Postle, 2015; Grill-Spector & Weiner, 2014) for face and scene stimuli) were most important for identifying PM intention maintenance (as identified by a surface-based searchlight analysis (Kriegeskorte, Goebel, & Bandettini, 2006; Oosterhof, Connolly, & Haxby, 2016); see **extended data Figure 3-2**). The searchlight results suggest that PM-item retrieval is primarily mediated by these more posterior regions, while the prefrontal region may have a more abstracted involvement, like PM-state or rule maintenance (i.e. maintaining whether or not one has a prospective intention at the moment). However, the lack of above chance PM-intention decoding in the PFC may be related to a lower signal-to-noise ratio or due to the "mixed selectivity" of prefrontal neurons (Bhandari, Gagne, & Badre, 2018). Further research is needed in order to better understand how and where PM intentions are represented and the specific role of classically identified regions like the anterior prefrontal cortex.

Our results are consistent with the dual mechanisms of cognitive control framework (Braver, 2012; Braver et al., 2007) and the dynamic multiprocess framework of PM (Scullin et al., 2013; Shelton & Scullin, 2017). Both dual-mechanism frameworks posit that individuals can use two different methods of cognitive control to fulfill prospective intentions, and that they can flexibly adjust their control strategy in response to environmental factors, such as cognitive load or PM target expectancy. However, neither framework formally describes whether that adjustment is a fluid process or all-or-none "switch" between strategies. Here, our evidence suggests that there are graded levels of control between proactive and reactive control that people engage along a continuum.

An alternative explanation for the present results is that shifts in PM cost may not reflect shifts between proactive and reactive control strategies per se, but rather shifts between stronger and weaker levels of proactive control. Unitary models of PM such as the preparatory attention and memory "PAM" theory (Smith, 2003, p. 200) propose that successfully fulfilling prospective intentions relies on some level of proactive preparation in all situations. However, we believe this interpretation of our data is less likely than the dual-mechanisms account. According to the PAM model, we should expect extremely poor PM performance when evidence of proactive control is absent, however this was not the case. As mentioned previously, performance on trials where there were no observed costs on end probes (n-3 to n-1 before PM- target), PM accuracy was still higher than the PAM model would predict (mean = 67.81%, SEM = 2.67%). PM performance was also strong on trials where the neural measure of PM intention evidence was absent (mean = 81.1%, SE = 3.2%, average N = 11.5 trials/participant), and also on trials where both PM cost and PM intention evidence were absent (mean = 70.5%, SE = 6.5%, average N = 2.8 trials/participant). Incidentally, the link between PM cost and PM accuracy in the present study closely replicates prior work using a similar, though static, task design (Lewis-Peacock et al., 2016). In that experiment, the researchers found that on blocks where participants were biased towards reactive control, PM accuracy was 66.0% (SEM: 4.1%), and when participants were biased towards proactive control, PM accuracy was 71.2% (SEM: 3.0%).

One limitation of the current study is its reliance on the behavioral PM cost measure to infer PM strategy use. While this has become a standard approach, PM cost is nonetheless an indirect measure of PM strategy, the underlying source of which is still

under debate (Ball et al., 2015; Boywitt & Rummel, 2012; Heathcote, Loft, & Remington, 2015; Strickland, Heathcote, Remington, & Loft, 2017). We sought to complement this indirect measure with a more direct measure of PM intention processing using fMRI pattern classifiers to track PM intention maintenance. However, fMRI is sluggish and not ideal to observe time-sensitive shifts in neural coding. It is possible that our neural measure incorporates aspects of both PM-intention maintenance as well as strategic monitoring. However, these neural measures were not correlated with PM cost slopes, which are believed to reflect changes in monitoring. In addition, previous work has found that in contexts where participants are biased towards reactive control, the level of monitoring is not related to PM performance (Harrison & Einstein, 2010; Loft, Bowden, Ball, & Brewer, 2014), however it is related to PM performance when participants are biased towards proactive control (Ball et al., 2015; Lewis-Peacock et al., 2016; Loft, Bowden, Ball, & Brewer, 2014).

In conclusion, we developed a novel dual-task paradigm to show that people solve prospective memory problems by flexibly shifting between proactive control and reactive control in response to changes in ongoing cognitive demands. We found evidence for two different components of proactive control – strategic monitoring, measured behaviorally, and PM intention maintenance, measured neurally – which independently fluctuate and contribute to PM performance. These shifts were adaptive in that adjustments of control towards the strategy favored for a given situation (e.g., shifting towards proactive control when demands decreased across time) led to better PM performance. These results extend dual mechanism accounts of PM by

demonstrating that cognitive flexibility (i.e., adapting cognitive control strategies to the environment) is beneficial for remembering to perform future intentions.

References

Anderson, F. T., McDaniel, M. A., & Einstein, G. O. (2017). Remembering to Remember: An Examination of the Cognitive Processes Underlying Prospective Memory ☆. In *Learning and Memory: A Comprehensive Reference* (pp. 451–463). https://doi.org/10.1016/B978-0-12-809324-5.21049-3

Avants, B. B., Tustison, N. J., Song, G., Cook, P. A., Klein, A., & Gee, J. C. (2011). A reproducible evaluation of ANTs similarity metric performance in brain image registration. *NeuroImage*, *54*(3), 2033–2044. https://doi.org/10.1016/j.neuroimage.2010.09.025

Ball, B. H., Brewer, G. A., Loft, S., & Bowden, V. (2015). Uncovering continuous and transient monitoring profiles in event-based prospective memory. *Psychonomic Bulletin & Review*, *22*(2), 492–499. https://doi.org/10.3758/s13423-014-0700-8

Ballhausen, N., Schnitzspahn, K. M., Horn, S. S., & Kliegel, M. (2017). The interplay of intention maintenance and cue monitoring in younger and older adults' prospective memory. *Memory & Cognition*, *45*(7), 1113–1125. https://doi.org/10.3758/s13421-017-0720-5

Beck, S. M., Ruge, H., Walser, M., & Goschke, T. (2014). The functional neuroanatomy of spontaneous retrieval and strategic monitoring of delayed intentions. *Neuropsychologia*, *52*, 37–50. https://doi.org/10.1016/j.neuropsychologia.2013.10.020

Bisiacchi, P. S., Schiff, S., Ciccola, A., & Kliegel, M. (2009). The role of dual-task and task-switch in prospective memory: Behavioural data and neural correlates. *Neuropsychologia*, *47*(5), 1362–1373. https://doi.org/10.1016/j.neuropsychologia.2009.01.034

Bowden, V., K., Smith, R. E., & Loft, S. (2017). Eye Movements Provide Insights into the Conscious Use of Context in Prospective Memory. *Consciousness and Cognition*, *52*, 68–74.

Boywitt, D., C., & Rummel, J. (2012). A diffusion model analysis of task interference effects in prospective memory. *Memory & Cognition*, *40*(1), 70–82.

Braver, T. S. (2012). The variable nature of cognitive control: a dual mechanisms framework. *Trends in Cognitive Sciences*, *16*(2), 106–113. https://doi.org/10.1016/j.tics.2011.12.010

Braver, T. S., Gray, J., R., & Burgess, Z. (2007). Explaining the many varieties of working memory variation: dual mechanisms of cognitive control. In *Variation in working memory* (pp. 142–214). Oxford University Press.

Brewer, G. A., Knight, J. B., Marsh, R. L., & Unsworth, N. (2010). Individual differences in event-based prospective memory: Evidence for multiple processes supporting cue detection. *Memory & Cognition*, *38*(3), 304–311. https://doi.org/10.3758/MC.38.3.304

Cohen, A.-L. (2008). Number of cues influences the cost of remembering to remember. *Memory & Cognition*, *36*(1), 149–156.

Cohen, A.-L., Gordon, A., Jaudas, A., Hefer, C., & Dreisbach, G. (2017). Let it go: the flexible engagement and disengagement of monitoring processes in a non-focal

prospective memory task. *Psychological Research*, *81*(2), 366–377. https://doi.org/10.1007/s00426-016-0744-7

Cona, G., Bisiacchi, P., S., & Moscovitch, M. (2014). The Effects of Focal and Nonfocal Cues on the Neural Correlates of Prospective Memory: Insights From ERPs. *Cerebral Cortex*, *24*(10), 2630–2646.

Cona, G., Scarpazza, C., Sartori, G., Moscovitch, M., & Bisiacchi, P. S. (2015). Neural bases of prospective memory: A meta-analysis and the "Attention to Delayed Intention" (AtoDI) model. *Neuroscience & Biobehavioral Reviews*, *52*, 21–37.

Crovitz, H. F., & Daniel, W. F. (1984). Measurements of everyday memory: Toward the prevention of forgetting. *Bulletin of the Psychonomic Society*, *22*(5), 413–414. https://doi.org/10.3758/BF03333861

D'Esposito, M., & Postle, B. R. (2015). The Cognitive Neuroscience of Working Memory. *Annual Review of Psychology*, *66*(1), 115–142. https://doi.org/10.1146/annurev-psych-010814-015031

Dismukes, R. K. (2012). Prospective Memory in Workplace and Everyday Situations. *Current Directions in Psychological Science*, *21*(4), 215–220. https://doi.org/10.1177/0963721412447621

Einstein, G. O., & McDaniel, M. A. (2005). Prospective Memory: Multiple Retrieval Processes. *Current Directions in Psychological Science*, *14*(6), 286–290. https://doi.org/10.1111/j.0963-7214.2005.00382.x

Einstein, G. O., & McDaniel, M. A. (2010). Prospective memory and what costs do not reveal about retrieval processes: A commentary on Smith, Hunt, McVay, and

McConnell (2007). *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*(4), 1082–1088. https://doi.org/10.1037/a0019184

Efron, B. (1979). Bootstrap Methods: Another Look at the Jackknife. *The Annals of Statistics*, 7(1), 1-26.

Fischl, B. (2012). FreeSurfer. *NeuroImage*, *62*(2), 774–781. https://doi.org/10.1016/j.neuroimage.2012.01.021

Gilbert, S., J. (2011). Decoding the content of delayed intentions. *The Journal of Neuroscience*, *31*(8), 2888–2894.

Gilbert, S. J., Hadjipavlou, N., & Raoelison, M. (2013). Automaticity and Control in Prospective Memory: A Computational Model. *PLoS ONE*, *8*(3), e59852. https://doi.org/10.1371/journal.pone.0059852

Gratton, G., Cooper, P., Fabiani, M., Carter, C. S., & Karayanidis, F. (2018). Dynamics of cognitive control: Theoretical bases, paradigms, and a view for the future. *Psychophysiology*, *55*(3), e13016. https://doi.org/10.1111/psyp.13016

Grill-Spector, K., & Weiner, K. S. (2014). The functional architecture of the ventral temporal cortex and its role in categorization. *Nature Reviews Neuroscience*, *15*(8), 536–548. https://doi.org/10.1038/nrn3747

Guynn, M. J. (2003). A two-process model of strategic monitoring in event-based prospective memory: Activation/retrieval mode and checking. *International Journal of Psychology*, *38*(4), 245–256. https://doi.org/10.1080/00207590344000178

Guynn, M., J. (2008). Theory of Monitoring in Prospective Memory Instantiating a Retrieval Mode and Periodic Target Checking. In *Prospective Memory: Cognitive,*

Neuroscience, Developmental, and Applied Perspectives (pp. 53–76). New York: Psychology Press.

Harrison, T. L., & Einstein, G. O. (2010). Prospective memory: Are preparatory attentional processes necessary for a single focal cue? *Memory & Cognition*, *38*(7), 860–867. https://doi.org/10.3758/MC.38.7.860

Harrison, T. L., Mullet, H. G., Whiffen, K. N., Ousterhout, H., & Einstein, G. O. (2014). Prospective memory: Effects of divided attention on spontaneous retrieval. *Memory & Cognition*, *42*(2), 212–224. https://doi.org/10.3758/s13421-013-0357-y

Heathcote, A., Loft, S., & Remington, R., W. (2015). Slow Down and Remember to Remember! A Delay Theory of Prospective Memory Costs. *Psychological Review*, *122*(2), 376–410.

Hicks, J. L., Marsh, R. L., & Cook, G. I. (2005). Task interference in time-based, event-based, and dual intention prospective memory conditions☆. *Journal of Memory and Language*, *53*(3), 430–444. https://doi.org/10.1016/j.jml.2005.04.001

Horn, S. S., Bayen, U. J., & Smith, R. E. (2013). Adult age differences in interference from a prospective-memory task: a diffusion model analysis. *Psychonomic Bulletin & Review*, *20*(6), 1266–1273. https://doi.org/10.3758/s13423-013-0451-y

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

Kim, H., Daselaar, S.M., Cabeza, R. (2010). Overlapping brain activity between episodic memory encoding and retrieval: Roles of the task-positive and task-negative networks. NeuroImage, 49, 1045-1054.

1021 Kiyonaga, A., Dowd, E.W., Egner, T. (2017). Neural representation of working memory
1022 content is modulated by visual attentional demand. *Journal of Cognitive*
1023 *Neuroscience*, 29(12), 2011-2024.
1024 Kliegel, M., & Martin, M. (2003). Prospective memory research: Why is it relevant?
1025 *International Journal of Psychology*, *38*(4), 193–194.
1026 Knight, J. B., Meeks, J. T., Marsh, R. L., Cook, G. I., Brewer, G. A., & Hicks, J. L.
1027 (2011). An observation on the spontaneous noticing of prospective memory
1028 event-based cues. *Journal of Experimental Psychology: Learning, Memory, and*
1029 *Cognition*, *37*(2), 298–307. https://doi.org/10.1037/a0021969
1030 Kriegeskorte, N., Goebel, R., & Bandettini, P. (2006). Information-based functional brain
1031 mapping. *Proceedings of the National Academy of Sciences*, *103*(10), 3863–
1032 3868. https://doi.org/10.1073/pnas.0600244103
1033 Kuhlmann, B. G., & Rummel, J. (2014). Context-specific prospective-memory
1034 processing: Evidence for flexible attention allocation adjustments after intention
1035 encoding. *Memory & Cognition*, *42*(6), 943–949. https://doi.org/10.3758/s13421-
1036 014-0405-2
1037 Lewis-Peacock, J. A., Cohen, J. D., & Norman, K. A. (2016). Neural evidence of the
1038 strategic choice between working memory and episodic memory in prospective
1039 remembering. *Neuropsychologia*, *93*, 280–288.
1040 https://doi.org/10.1016/j.neuropsychologia.2016.11.006
1041 Li, X., Morgan, P. S., Ashburner, J., Smith, J., & Rorden, C. (2016). The first step for
1042 neuroimaging data analysis: DICOM to NIfTI conversion. *Journal of*

Neuroscience Methods, *264*, 47–56. https://doi.org/10.1016/j.jneumeth.2016.03.001

Loft, S., Bowden, V. K., Ball, B. H., & Brewer, G. A. (2014). Fitting an ex-Gaussian function to examine costs in event-based prospective memory: Evidence for a continuous monitoring profile. *Acta Psychologica*, *152*, 177–182. https://doi.org/10.1016/j.actpsy.2014.08.010

Marklund, P., & Persson, J. (2012). Context-dependent switching between proactive and reactive working memory control mechanisms in the right inferior frontal gyrus. *NeuroImage*, *63*(3), 1552–1560. https://doi.org/10.1016/j.neuroimage.2012.08.016

McDaniel, M. A., & Einstein, G. O. (2000). Strategic and automatic processes in prospective memory retrieval: a multiprocess framework. *Applied Cognitive Psychology*, *14*(7), S127–S144. https://doi.org/10.1002/acp.775

McDaniel, M., A., & Einstein, G., O. (2007). *Prospective memory: an overview and synthesis of an emerging field*. Los Angeles: SAGE Publications.

McDaniel, M. A., LaMontagne, P., Beck, S. M., Scullin, M. K., & Braver, T. S. (2013). Dissociable Neural Routes to Successful Prospective Memory. *Psychological Science*, *24*(9), 1791–1800. https://doi.org/10.1177/0956797613481233

Momennejad, I., & Haynes, J.-D. (2012). Human anterior prefrontal cortex encodes the 'what' and 'when' of future intentions. *NeuroImage*, *61*(1), 139–148. https://doi.org/10.1016/j.neuroimage.2012.02.079

Momennejad, I., & Haynes, J.-D. (2013). Encoding of Prospective Tasks in the Human Prefrontal Cortex under Varying Task Loads. *The Journal of Neuroscience*, *33*(44), 17342–17349.

Oksanen, K. M., Waldum, E. R., McDaniel, M. A., & Braver, T. S. (2014). Neural Mechanisms of Time-Based Prospective Memory: Evidence for Transient Monitoring. *PLoS ONE*, *9*(3), e92123. https://doi.org/10.1371/journal.pone.0092123

Oosterhof, N. N., Connolly, A. C., & Haxby, J. V. (2016). CoSMoMVPA: Multi-Modal Multivariate Pattern Analysis of Neuroimaging Data in Matlab/GNU Octave. *Frontiers in Neuroinformatics*, *10*. https://doi.org/10.3389/fninf.2016.00027

R Core Team (2017). R: A language and environment for statistical computing. R. Foundation for Statistical Computing, Vienna, Austria. URL: https://www.R-project.org

Reynolds, J. R., West, R., & Braver, T. (2009). Distinct Neural Circuits Support Transient and Sustained Processes in Prospective Memory and Working Memory. *Cerebral Cortex*, *19*(5), 1208–1221. https://doi.org/10.1093/cercor/bhn164

Rummel, J., & Meiser, T. (2013). The role of metacognition in prospective memory: Anticipated task demands influence attention allocation strategies. *Consciousness and Cognition*, *22*(3), 931–943. https://doi.org/10.1016/j.concog.2013.06.006

Scullin, M. K., Bugg, J. M., & McDaniel, M. A. (2012). Whoops, I did it again: Commission errors in prospective memory. *Psychology and Aging*, *27*(1), 46–53. https://doi.org/10.1037/a0026112

Scullin, M. K., McDaniel, M. A., & Einstein, G. O. (2010). Control of cost in prospective memory: Evidence for spontaneous retrieval processes. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*(1), 190–203. https://doi.org/10.1037/a0017732

Scullin, M. K., McDaniel, M. A., & Shelton, J. T. (2013). The Dynamic Multiprocess Framework: Evidence from prospective memory with contextual variability. *Cognitive Psychology*, *67*(1–2), 55–71. https://doi.org/10.1016/j.cogpsych.2013.07.001

Scullin, M. K., McDaniel, M. A., Shelton, J. T., & Lee, J. H. (2010). Focal/nonfocal cue effects in prospective memory: Monitoring difficulty or different retrieval processes? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*(3), 736–749. https://doi.org/10.1037/a0018971

Shelton, J. T., & Scullin, M. K. (2017). The Dynamic Interplay Between Bottom-Up and Top-Down Processes Supporting Prospective Remembering. *Current Directions in Psychological Science*, *26*(4), 352–358. https://doi.org/10.1177/0963721417700504

Smith, R. E. (2003). The cost of remembering to remember in event-based prospective memory: Investigating the capacity demands of delayed intention performance. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*(3), 347–361. https://doi.org/10.1037/0278-7393.29.3.347

Smith, R. E., Hunt, R. R., McVay, J. C., & McConnell, M. D. (2007). The cost of event-based prospective memory: Salient target events. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*(4), 734–746. https://doi.org/10.1037/0278-7393.33.4.734

Sobel, K. V., Gerrie, M. P., Poole, B. J., & Kane, M. J. (2007). Individual differences in working memory capacity and visual search: The roles of top-down and bottom-up processing. *Psychonomic Bulletin & Review*, *14*(5), 840–845. https://doi.org/10.3758/BF03194109

Strickland, L., Heathcote, A., Remington, R. W., & Loft, S. (2017). Accumulating evidence about what prospective memory costs actually reveal. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *43*(10), 1616–1629. https://doi.org/10.1037/xlm0000400

Thompson, B. (1993). The Use of Statistical Significance Tests in Research: Bootstrap and Other Alternatives. *The Journal of Experimental Education*, *61*(4), 361-377.

Underwood, A. G., Guynn, M. J., & Cohen, A.-L. (2015). The Future Orientation of Past Memory: The Role of BA 10 in Prospective and Retrospective Retrieval Modes. *Frontiers in Human Neuroscience*, *9*, 1–12. https://doi.org/10.3389/fnhum.2015.00668

Wagenmakers, E.-J., & Farrell, S. (2004). AIC model selection using Akaike weights. *Psychonomic Bulletin & Review*, *11*(1), 192–196. https://doi.org/10.3758/BF03206482

West, R., Bowry, R., & Krompinger, J. (2006). The effects of working memory demands on the neural correlates of prospective memory. *Neuropsychologia*, *44*(2), 197–207. https://doi.org/10.1016/j.neuropsychologia.2005.05.003

West, R., Carlson, L., & Cohen, A.-L. (2007). Eye movements and prospective memory: What the eyes can tell us about prospective memory. *International Journal of Psychophysiology*, *64*, 269–277.

West, R., McNerney, M. W., & Travers, S. (2007). Gone but not forgotten: The effects of cancelled intentions on the neural correlates of prospective memory. *International Journal of Psychophysiology*, *64*(3), 215–225. https://doi.org/10.1016/j.ijpsycho.2006.09.004

Whitehead, P., S., & Egner, T. (2018). Cognitive Control Over Prospective Task-Set Interference. *Journal of Experimental Psychology: Human Perception and Performance*, *44*(5), 741–755.

Zuber, S., Kliegel, M., & Ihle, A. (2016). An individual difference perspective on focal versus nonfocal prospective memory. *Memory & Cognition*, *44*(8), 1192–1203. https://doi.org/10.3758/s13421-016-0628-5

Figure 1-1. Average reaction time (y-axis left, blue) and accuracy (y-axis right, red) for pilot participants are plotted across each difficulty level of the ongoing visual search task. Reaction time increases and accuracy decreases from the easiest difficulty (1) to the hardest (14). The purpose of the first behavioral pilot study was to determine if the ongoing task could be parametrically modulated in a controlled manner. For this pilot study, participants (n = 15) performed the ongoing task in isolation. On each probe, participants indicated the absence or presence of the arrow target (rightward facing horizontal arrow) in a newly generated visual-search array (every 2 s) with a button press (left: absent; right: present; response deadline: 1.9 s). Target arrow location was counterbalanced between the top and bottom half of the screen. Non-target arrows appeared in set positions around the circular array, oriented within some distribution of angles determined by the current task difficulty setting. Participants sat approximately 18 in. away from the screen, and all 10 arrows, which were .64° by .22° in shape, were 3.18° away from the center of the screen. OG task difficulty was manipulated on each probe by adjusting two parameters controlling the orientation of the distractor arrows: their minimum similarity to the target and their similarity to other distractors. For distractor-to-target similarity, a minimum angular distance for distractors from the target (i.e., horizontal or 0°) was set to either 5, 15, 25, 45, or 65 degrees. For distractor-to-distractor similarity, the maximum variance from the minimum angular distance was set to either 10, 20, or 40 degrees. The factorial combination of these parameters (excluding any combination where minimum plus variance could exceed the 90° vertical plane) created 14 difficulty conditions. Participants performed three blocks of trials separated by short voluntary breaks for rest. Each block contained 14 mini-blocks comprised of 20 visual-search trials of one difficulty level. Difficulty level was pseudo-randomly selected, with the only limitation being that each of the 14 difficulty levels was selected exactly three times. As difficulty increased, accuracy decreased ($F(13,182)=74.89$, $p<.001$) and reaction time increased ($F(13,182)=39.53$, $p<.001$).

Figure 1-2. The purpose of the second behavioral pilot study was to determine if PM strategy (as measured by PM cost) could be modulated by the difficulty of the ongoing task. The task design for this study was nearly identical to that used in the main experiment, but here OG task difficulty was held constant as either *easy* (difficulty level 4) or *hard* (difficulty level 12) for the entirety of each block and across each trial. Participants (n = 20) completed one block (15 trials per block) at each difficulty level. **a)** PM cost was calculated by subtracting the average non-PM trial OG RT from the average PM trial OG RT at each difficulty level for each participant. The distribution and median are represented. PM cost significantly varied as a function of OG task difficulty ($F(1,19) = 35.63$, $p<.001$), with cost being higher at the easier difficulty (M = 0.134 s (SE = 0.012)) than at the harder difficulty (M = 0.031 s (SE = 0.012)). * $p<.05$. **b)** PM accuracy was calculated at each difficulty for each participant. The distribution and median are represented. PM accuracy was equivalent across conditions ($F(1,19) = 0.785$, $p = 0.387$; easy = 71.0% (4.5%), hard = 64.5% (5.8%)).

Figure 2-1. Additional model comparisons for trial-by-trial estimates of PM strategy shifts. One concern with fitting models on a by-trial basis is that noise may bias model selection towards the models with fewer parameters. In order to address this concern, we performed a less conservative AIC (without the small sample correction term) model selection analysis and a bootstrap analysis where polynomial fits were calculated for a random subsample of trials. For the AIC analysis, we performed the same regression analysis steps as detailed for the AICc analysis, but simply used the AIC estimation term instead of the AICc term. After calculating an AIC score for each trial, we then selected the lowest score between the 1st, 2nd, and 3rd order polynomial as the best for the trial. We then calculated the relative Akaike weight for each model on each trial and average that value for each model type for each participant. Across participants the average Akaike weights were similar between 1st and 3rd order polynomial fits (mean difference = 0.6 (SE = 1.8), t(77) = 0.65, p=.52; **panel a**). Significantly more trials for each participant were best fit by a 1st order than a 3rd order polynomial (mean difference = 20.5% (SE = 3.4%), t(77) = 12.014, p<.001; **panel b**). Another way to mitigate the impact of noise on model selection is to fit the model on more than a single trial at a time. To avoid averaging across all trials and still getting an estimate of model-fit reliability, we performed a bootstrap analysis. In this analysis, we first z-scored PM-cost values within each subject. Next, we combined all trials into one super-subject. On each bootstrap iteration, 50 trials of each of the five trial types (increasing starting easy, increasing starting middle, fixed, decreasing starting hard, decreasing starting easy) were randomly selected from the super-subject pool. Then, 1st, 2nd, and 3rd order polynomial models were fit to each trial type sample and the lowest AIC value was selected as the best-fit model type. We repeated this process 1000 times and found that a linear (1st order) polynomial fit a significantly greater number of these samples (57.7% of all trials), followed by a quadratic fit (2nd order, 29.1% of all trials), followed by a cubic fit (3rd order, 13.4% of all trials). This is a significantly greater number of trials selected to be best fit by a linear relationship than would be predicted by chance (χ^2 (1, n=1000) = 364.06, p <.001).

A		**Behavioral Sample**	**Neural Sample**	**Combined Sample**
Ongoing Task Accuracy		**(N=50)**	**(N=28)**	**(N=78)**
Overall	Mean	81.26%	83.02%	83.64%
	95% CI	[69.52%,93.00%]	[75.08%, 90.97%]	[75.49%,91.79%]
Difficulty (1 to 15)	β	-0.024	-0.030	-0.026
	95% CI	[-0.025, -0.023]	[-0.031, -0.028]	[-0.027, -0.025]
	p	<.001**	<.001**	<.001**
Trial Type (PM, non-PM)	β	0.004	0.017	0.008
	95% CI	[-0.005, 0.012]	[0.003, 0.030]	[0.001, 0.016]
	p	0.63	0.013*	0.024*
Difficulty x Trial Type	β	$-9.4*10^{-5}$	0.002	$6.2*10^{-4}$
	95% CI	[-0.002, 0.002]	[-0.001, 0.005]	[-0.001, 0.002]
	p	0.925	0.224	0.468
Main Effect Model	marginal R^2	0.552	0.604	0.569
B				
Ongoing Task RT				
Overall	Mean	1.089	1.113	1.098
	95% CI	[0.895, 1.284]	[0.857, 1.369]	[0.880, 1.316]
Difficulty (1 to 15)	β	0.034	0.033	0.033
	95% CI	[0.031, 0.036]	[0.030, 0.037]	[0.031, 0.036]
	p	< .001**	< .001**	< .001**
Trial Type (PM, non-PM)	β	-0.102	-0.139	-0.115
	95% CI	[-0.111, -0.092]	[-0.153, -0.125]	[-0.123, -0.107]
	p	< .001**	< .001**	< .001**
Difficulty x Trial Type	β	0.010	0.014	0.011
	95% CI	[0.008, 0.012]	[0.011, 0.017]	[0.010, 0.013]
	p	< .001**	< .001**	< .001**
Interaction Model	marginal R^2	0.553	0.486	0.525
C				
PM Cost				
Difficulty	β	-8.217	-14.167	-10.353

(1 to 15)	95% CI	[-10.471, -5.964]	[-17.352, -10.982]	[-12.289, -8.417]
	p	< .001**	< .001**	< .001**
D				
PM Cost Slope				
Trial Type	F	F(2,98) = 22.39	F(2,54) = 28.53	F(2,154) = 47.02
(Dec, Fix, Inc)	p	< .001**	< .001**	< .001**
	Mean	4.78	8.28	6.03
Decreasing	95% CI	[3.41, 6.14]	[6.25, 10.31]	[4.86, 7.21]
	p(mean>0)	< .001**	< .001**	< .001**
	Mean	3.23	3.53	3.34
Fixed	95% CI	[1.59,4.87]	[1.43, 5.64]	[2.07, 4.61]
	p(mean>0)	< .001**	0.002*	< .001**
	Mean	-1.03	-1.33	-1.14
Increasing	95% CI	[-2.34, 0.28]	[-3.42,0.76]	[-2.24, -0.42]
	p(mean<0)	0.120	0.202	0.042*
Increasing vs. Fixed	ANOVA	F(1,49) = 20.16	F(1,27) = 17.38	F(1,77) = 37.14
	p	< .001**	< .001**	< .001 **
	ANOVA	F(1,49) = 3.32	F(1,27) = 18.13	F(1,77) = 15.10
Decreasing vs. Fixed	p	0.075	< .001**	< .001**
Increasing vs. Decreasing	ANOVA	F(1,49) = 41.99	F(1,27) = 41.02	F(1,77) = 78.63
	p	< .001**	< .001**	< .001**
E				
PM Accuracy				
Overall	Accuracy	69.66%	83.33%	74.57%
	95% CI	[50.35%,88.97%]	[73.48%, 93.17%]	[56.82%, 92.31%]
Trial Type	ANOVA	F(2,98) = 1.19	F(2,54) = 0.36	F(2,154) = 0.68
	p	0.310	0.699	0.508
	Accuracy	70.86%	82.32%	74.97%
Decreasing	95% CI	[64.78%, 76.94%]	[77.92%, 86.71%]	[70.65%, 79.29%]
	p(mean=0)	< .001**	< .001**	< .001**

Fixed	Accuracy	67.91%	83.00%	73.32%
	95% CI	[61.67%, 74.14%]	[76.49%, 89.49%]	[68.50%, 78.14%]
	p(mean=0)	< .001**	< .001**	< .001**
Increasing	Accuracy	69.39%	84.61%	74.86%
	95% CI	[63.99%, 74.80%]	[80.28%, 88.95%	[70.77%, 78.94%]
	p(mean=0)	< .001**	< .001**]	< .001**
False Alarm Rate	% of probes	0. 71% (SE = 0.11%)	0.41% (SE = 0.05%)	0.60% (SE = .01%)
F				
PM Accuracy: Trial Direction * PM Cost Slope				
Decreasing	β	0.013	0.026	0.018
	95% CI =	[0.003, 0.025]	[0.008, 0.047]	[0.008, 0.028]
	p	0.007*	0.003*	< .001**
Fixed	β	0.018	0.025	0.020
	95% CI =	[0.003, 0.034]	[-0.003, 0.058]	[0.006, 0.033]
	p	0.011*	0.041*	0.003*
Increasing	β	-0.006	-0.003	-0.006
	95% CI =	[-0.016, 0.004]	[-0.025, 0.018]	[-0.015, 0.004]
	p	0.126	0.401	0.119
Decreasing vs. Increasing	β	0.019	0.0290	0.024
	95% CI =	[0.005, 0.034]	[-0.002, 0.061]	[0.011, 0.037]
	p	0.006*	0.031*	< .001 **
Decreasing vs. Fixed	β	0.005	-0.001	0.001
	95% CI =	[-0.013, 0.022]	[-0.035, 0.035]	[-0.015, 0.017]
	p	0.305	0.534	0.423
Increasing vs. Fixed	β	0.024	0.028	0.025
	95% CI =	[0.005, 0.0421]	[-0.003, 0.061]	[0.009, 0.041]
	p	0.007*	0.040*	0.025*

* p < 0.05, ** p < 0.001

Figure 2-2. Comparison of behavioral results from experiments 1 and 2. These data include the key results presented in **Figure 2** separately for the behavioral-only participants, the neural participants, and the combined groups. Each analysis section of the table (A-F) corresponds to the same panel from **Figure 2**. Analyses of the relationship between trial type (PM/non-PM), Difficulty (1 to 15) and OG task accuracy, OG RT, and PM cost were carried out by first running a mixed-effect regression using the *lme4* package in R of the interaction between trial type and difficulty and then a separate model comparing the main effects of difficulty and trial type without the interaction term. Random effects of individual slope and intercept were included in each regression. Within-subject ANOVAs were used to compare PM cost slope and PM accuracy across conditions.

Figure 2-3. Visualization of model comparisons for by-trial estimates of PM strategy shifts. Here, we used AICc to evaluate the relative model fit between a linear, quadratic, and cubic relationship of PM cost over the course of each trial. The Akaike weight calculated from AICc scores from each model are shown in **Figure 2-3a**. We evaluated the lowest AICc for each trial, and then calculated the proportion of trials best fit by either a linear, quadratic, and cubic relationship for each participant (shown in **Figure 2-3b**). A 1st order polynomial (linear model) fit best for nearly all trials (mean = 93.43%, 95% CI = [92.62%, 94.24%]; Akaike weight = 0.873, 95% CI = [0.865, 0.880]), compared to 2nd order fits (mean = 5.72%, 95% CI = [4.99%, 6.45%]; Akaike weight = 0.111, 95% CI = [0.104, 0.119]), or 3rd order fits (mean = 0.85%, 95% CI = [0.68%, 1.02%]; Akaike weight = 0.016, 95% CI = [0.012, 0.019]).

Cluster Index	Z	x	y	z	ROI Label	Hemisphere
5	8.07	10	-98	6	Occipital Pole	R
5	7.83	-26	-58	-8	Fusiform	L
5	7.5	32	-56	-14	Fusiform	R
5	7.36	-4	-72	-24	Cerebellum	L
5	7.24	-18	-62	8	Intracalcarine Cortex	L
5	7.2	6	-70	-26	Cerebellum	R
4	4.98	36	36	32	dlPFC (BA 9)	R
4	4.95	34	54	14	rlPFC (BA10)	R
4	4.84	24	56	8	rlPFC (BA10)	R
4	4.75	36	42	34	dlPFC (BA 9)	R
4	4.75	38	52	14	rlPFC (BA10)	R
4	4.74	28	60	8	rlPFC (BA10)	R
3	5.85	34	22	10	anterior insula	R
3	5.26	32	28	-2	anterior insula	R
3	5.24	32	26	2	anterior insula	R
3	5.19	36	28	10	anterior insula	R
3	5.18	42	18	6	anterior insula	R
3	4.99	38	16	24	anterior insula	R
2	5.51	-8	-38	24	posterior cingulate	L
2	5.44	8	-22	26	posterior cingulate	R
2	5.31	-8	-24	26	posterior cingulate	L
2	5.25	-6	-28	26	posterior cingulate	L
2	5.17	-8	-44	20	posterior cingulate	L
2	5.13	-6	-18	28	posterior cingulate	L
1	4.75	-30	50	10	rlPFC (BA10)	L
1	4.71	-26	42	6	rlPFC (BA10)	L
1	4.43	-34	58	14	rlPFC (BA10)	L
1	4.35	-36	54	20	rlPFC (BA10)	L
1	4.05	-34	58	20	rlPFC (BA10)	L
1	3.59	-22	50	-4	rlPFC (BA10)	L

Figure 3-1: PM > Non-PM GLM Analysis. FSL's FEAT was used to identify voxels that were more responsive on PM trials than on non-PM trials (cluster correction, $p < .001$). This table lists the coordinates and descriptors for all significant voxel clusters.

Figure 3-2. Surface searchlight analysis. Results from the surface-based searchlight classification analysis to decode the PM intention on PM trials. Vertices in red indicate those that survived threshold-free cluster enhancement significance testing (H_0 mean = 50%, $p<.001$). Results indicate that classification was successful only in two particular posterior regions: the ventral temporal cortex and lateral occipital cortex. Notably absent from this map is the anterior lateral prefrontal cortex. To perform this analysis, anatomical surface outputs from Freesurfer *recon-all* were converted to AFNI/SUMA format using *SUMA_Make_Spec_FS*. Surfaces were remapped to a standard topology using *MapIcosahedron* and co-registered to a reference functional volume using *align_epi_anat* so that functional data could be masked by the surface volume. Voxels determined to not be part of the surface were masked out of the searchlight analysis. Surface searchlight analysis was performed in MATLAB using functions from the CosMoMVPA toolbox. Each searchlight sphere was determined by selecting the 100 closest vertices to a center vertex according to geodesic distance. L2-weighted logistic regression classifiers were trained on four categories and tested within each searchlight sphere using a k-fold cross validation procedure. Only the five main PM task blocks were used for this analysis, and data from the localizers were excluded. That meant that on each k-fold iteration, 4 out of 5 PM-task blocks were used for training the classifier, and one held out block was used for testing. Accuracy across all five folds was averaged and that value was assigned to the center vertex of that sphere.