Human eye movements track stimulus meaning

Ali Pexsa[1], Edward F. Ester[1,2*]

[1]*Deparment of Psychology and Cognitive and Brain Sciences Graduate Program, Department of Psychology, University of Nevada, Reno*
[2]*Institute for Neuroscience, University of Nevada, Reno*

* Correspondence: eester@unr.edu

Author Contributions (CRedIT):

Conceptualization:	EFE, AP
Methodology:	EFE, AP
Software:	EFE
Validation:	AP, EFE
Formal Analysis:	EFE, AP
Investigation:	AP
Resources:	EFE
Data Curation:	EFE, AP
Writing – Original	EFE
Writing – Editing	AP, EFE
Visualization:	AP, EFE
Supervision:	EFE
Project Administration	EFE
Funding Acquisition	EFE

Support: NSF 2050833 (EFE)

Acknowledgements: We thank Ms. Jasdeep Buttar for assistance with data collection.

Abstract (142/150 words)

Working memory (WM) is a capacity- and duration-limited system that enables the representation and processing of fleeting sensory phenomena. Human fMRI studies indicate that brain areas associated with oculomotor control contribute to WM, while human eye tracking studies indicate that the physical characteristics of items held in WM can be decoded from eye movements. Whether eye movements also encode the semantic attributes of stimuli is unknown. Across three experiments, we demonstrate that stimulus meaning – operationalized via category membership – can be decoded from human gaze behavior (total N = 79 adults). Category decoding was driven by a mixture of eye movement parameters, could not be explained as a by-product of manual actions, and disappeared when participants attentively processed but did not categorize stimuli. These findings provide important and novel evidence demonstrating that human oculomotor behavior varies with the semantic meaning of a stimulus.

Statement of Relevance (136/150 words)

Working memory forms a temporal bridge between recent events and possible actions. Information stored in WM guides what we need to do next: where to look, where to grasp, or what to say. Recent studies have shown that human gaze behavior co-varies with the physical attributes of visual stimuli held in WM, for example, their locations and orientations. In naturalistic settings, actions are determined by what a stimulus means. We report that stimulus category – i.e., meaning – can be decoded from subtle variations in humans' eye movements while they categorized stimuli according to a remembered rule. This result disappeared when volunteers attended but did not categorize stimuli and could not be explained as a by-product of manual actions. These results reveal a behavioral "tell", where stimulus meaning can be extracted from subtle variations in where humans look.

Working memory (WM) is a capacity- and duration-limited system that forms a temporal bridge between sensory phenomena and possible actions. This system is a cornerstone of human executive function, as evidenced by robust correlations between measurements of WM performance and general cognitive aptitude (e.g., IQ; Cowan, 2001). Human fMRI studies indicate that visual WM storage depends, in part, on retinotopically organized brain areas (e.g., V1-V3 and IPS0-3; Harrison & Tong, 2009; Serences et al., 2009; Ester et al., 2009; Ester et al., 2013; Sprague et al., 2014; Sprague et al., 2016; Kwak & Curtis, 2022), including those implicated in oculomotor control, for example, the superior precentral sulcus (sPCS; regarded as the putative homologue of the monkey frontal eye fields; e.g., Ester et al., 2015; Shao et al., 2025).

Recent human studies have documented meaningful patterns of eye movements during WM (Mostert et al., 2018; Quax et al., 2019; van Ede et al., 2019; Linde-Domingo & Spitzer, 2023). For example, van Ede et al. (2019) required human volunteers to recall the orientations of extrafoveally presented stimuli by adjusting a foveal report stimulus and reported small (< 0.5°) but robust eye movements towards the location where a to-be-recalled stimulus previously appeared, even though the task provided no incentive for eye movements. Separately, Linde-Domingo and Spitzer (2024) required participants to remember the orientations of real-world objects presented at fovea and found that small gaze shifts (< 2.0°) robustly encoded the orientations of remembered items. Note that in both cases, gaze patterns co-varied with the physical attributes of remembered stimuli: their locations (van Ede et al., 2019) or their orientations (Linde-Domingo & Spitzer, 2024). Whether gaze patterns co-vary with stimulus meaning is unknown.

One way to study meaning is through categorization, defined as the process of assigning unique and behaviorally relevant labels to stimuli. Many categories are defined by rules: a red traffic light means “stop”, while a green traffic light means “go”. To study rule-based categorization, researchers train agents (i.e., animals or humans) to classify continuously varying stimuli into discrete groups based on an arbitrary rule. This type of learning has been associated with category-specific neural responses in several brain areas linked to WM storage, including V1-V3 (Ester et al., 2020; Henderson et al., 2025), and the prefrontal cortex (PFC; Freedman et al., 2001). Recent non-human primate studies have established causal relationships between rule-based categorization performance and neural activity in two brain areas associated with oculomotor control, the lateral intraparietal area (LIP; Zhou & Freedman, 2019; Zhou et al., 2023) and the superior colliculus (SC; Peysakovich et al., 2024). Additionally, a recent study that analyzed the gaze behavior of several monkeys trained to perform rule-based categorization tasks demonstrated that stimulus category could be decoded from eye movements (Rosen & Freedman, 2025). Notably, the experiments reviewed in this study required lengthy training regimens (tens- to hundreds-of-thousands of trials) and presented stimuli extrafoveally, i.e., in the spatial receptive fields of target SC and LIP neurons. Thus, the category-specific patterns of eye movements reported in this study could be influenced by task parameters or training histories.

Here, we tested the extent to which stimulus meaning – operationalized via category membership – could be predicted from eye movements in minimally trained human volunteers performing a rule-based categorization task at fovea. We found that stimulus category could be modestly but robustly decoded from eye position

(Experiment 1), that these effects were not a by-product of response planning or motor responses (Experiment 2), and that these effects were absent when participants attentively processed but did not categorize stimuli (Experiment 3). These findings demonstrate that human eye movements encode stimulus meaning, and they establish an important correspondence between oculomotor behaviors across species.

Methods

General Information Common to All Experiments

Participants. A total of 79 human volunteers (both sexes) enrolled in the 3 experiments reported here. 34 volunteers enrolled in Experiment 1, 27 volunteers enrolled in Experiment 2, and 16 volunteers enrolled in Experiment 3. Each volunteer was recruited from the local university community, aged 18-40, self-reported normal or corrected-to-normal visual acuity and participated in a single 1.5-hour testing session in exchange for course credit or monetary remuneration ($15/h). We had no a priori reason to expect that task performance or oculomotor behavior would vary as a function of race, ethnicity, or any other immutable characteristic; therefore, to better protect volunteers' privacy we did not collect this information.

Testing Environment and Stimuli. Stimuli were generated in MATLAB and rendered on a 27-inch g-sync enabled LCD monitor cycling at 240 Hz via Psychophysics Toolbox software extensions. The monitor resolution was fixed at 1920 x 1080 pixels. Participants were seated 85 cm from the display, with head position fixed by an SR Research chin rest. Under these circumstances, 1 degree of visual angle (DVA) corresponds to ~49.38 pixels.

Stimuli in each experiment were apertures containing 350 iso-oriented bars with inner and outer radii of 2 and 9 DVA, respectively. Stimuli flickered at 30 Hz (50% duty cycle), and each bar in the aperture was randomly replotted at the beginning of each "up" cycle to discourage volunteers from foveating individual bars. Each bar subtended 1 DVA with a stroke width of 6 pixels. Participants reported category judgments by pressing keys on a standard USB-powered keyboard.

Training Task. For all three experiments, volunteers learned to classify orientation stimuli into discrete groups according to an arbitrary (i.e., experimenter-chosen) boundary by completing a training task. Volunteers viewed apertures of iso-oriented bars and reported whether they were members of "Category 1" or "Category 2" by pressing the "Z" or "?" key (respectively) at any point during each 3 second trial (Figure S1A). Category membership was defined by a boundary whose orientation was randomly and independently selected for each participant on the interval [0°, 179°). We defined a set of stimuli tilted ±15°, ±30°, ±45°, and ±60° relative to each participant's category boundary and assigned stimuli tilted counterclockwise to the boundary to Category 1 and stimuli tilted clockwise to the boundary to Category 2 (Figure S1B and 1A). Feedback – i.e., "correct!" vs. "incorrect!" was given at the end of each trial. Failures to respond within the 3-second stimulus window were treated as incorrect. Participants received verbal instructions describing key features of the task. They were told that there existed an (unknown) boundary that divided stimuli into different categories, and that their task was to identify this boundary through trial-and-error. In general, with the benefit of these instructions, volunteers reached asymptotic performance by the 2nd or 3rd block, e.g., Figure S1C-E). Each volunteer completed 3 blocks of 32 trials in this task. Participants then completed one of the three experiments described below using the same (bespoke) category boundary imposed during training.

Due to a coding error, participants in Experiment 2 were trained using stimuli tilted ±15°, ±30°, ±45°, ±60°, and ±75° from the category boundary (instead of ±15°, ±30°, ±45°, and ±60°), but all other aspects of the task were identical to the training regimens used in Experiments 1 and 3.

Eyetracker Acquisition and Preprocessing. Binocular gaze position data was recorded via an SR Research Eyelink 1000 Plus infrared eyetracking system. Eyetracker recordings were obtained after applying a 9-point calibration procedure, and a drift correction procedure was applied at the beginning of each experimental block. Re-calibration was performed when necessary, for example, if the volunteer left the testing environment to use the restroom. Recordings were digitized at 500 Hz or 1 kHz.

Gaze data were minimally preprocessed. Data were downsampled from 1 kHz to 500 Hz when necessary and filtered for blinks. Blinks were defined as periods where the recorded pupil size fell below the median pupil size recorded across all experimental trials and conditions. We additionally discarded data from 50 ms before and 50 ms after each blink. Missing data were linearly interpolated, and the resulting gaze timeseries were smoothed with a gaussian kernel (10 ms full-width half-maximum). Saccades were identified using a velocity-based thresholding algorithm described in other work (e.g., Liu et al., 2024). Briefly, gaze position recordings were transformed into a velocity time course by calculating the Euclidian difference between successive samples. Saccades were identified as periods when gaze velocity exceeded 150 DVA. All findings reported here generalized across saccade thresholds ranging from 30 to 300 DVA.

Classification of Categories from Gaze Position. We used a decoding approach to determine the extent to which records of participants [x,y] gaze position differentiated between stimulus categories.

For time-resolved decoding (e.g., Figures 2A-B, 4C, 5D), we labeled trial-wise records of participants' gaze position according to category membership (i.e., Category 1 vs. Category 2). We implemented a leave-one-block-out cross-validation procedure, where gaze position records from N-1 blocks were used to train a linear discriminant classifier (via MATLAB's "classify" function) and the held-out block was used for testing. We used a linear discriminant function to minimize computational overhead; since decoding was performed over time, a support vector machine-based classifier trained took ~24 hours to complete for one subject, even with 32x parallelization. This procedure was repeated until all experimental blocks had been used as the test data set, and results were averaged across permutations. Decoding was performed on a timepoint-by-timepoint basis using minimally preprocessed data (see above).

For epoch-based decoding (e.g., Figures 2C-D, 4D, and 5E), we first extracted stable fixations using the velocity-based threshold procedure described in the preceding section. Fixations were labeled according to category membership and used to construct training and test data sets containing 80% and 20% of all fixations within an epoch of interest (e.g., sample period, delay period, etc.), respectively. The training data set was used to condition a support vector machine classifier (MATLAB's "fitcsvm" function), and the trained classifier was used to predict a category label for each observation in the test data set (MATLAB's "predict" function). We repeated this procedure 20 times while choosing different (randomly selected) subsets of data to

define the training and test data sets and averaged the results over permutations. Epoch-based decoding was computed separately using fixations recorded within the outer perimeter of the stimulus aperture (9 DVA; Figures 2, 4, and 5) or fixations recorded inside the inner perimeter of the stimulus aperture (i.e., microsaccades; 2 DVA; Figures S2, S4, and S5). Generally, this had little effect on decoding accuracy.

During Experiment 1, we omitted ambiguous trials (i.e., those where the presented orientation matched the participant's category boundary) from category decoding analyses. Data from these trials, however, were used to test the possibility that successful category decoding could be related to response demands (for example, looking at the keyboard key or response hand needed to produce the appropriate response on a given trial). To test that possibility, we trained a linear discriminant classifier to distinguish between categories based on records of gaze position measured during unambiguous trials, then applied the trained classifier to records of participants choice behavior (i.e., Category 1 or Category 2) during ambiguous trials. We reasoned that if decoding accuracy was driven by response selection or motor activity, then the trained classifier should be able to predict participants' responses during ambiguous trials when no category information was available.

Statistical Tests. Statistical analyses of time-resolved decoding accuracy (e.g., Figures 2A-B, Figure 4C, and Figure 5 D and F) were based on nonparametric signed randomization tests (Maris & Oostenveld, 2007). Unless otherwise specified, each test we performed assumes a null statistic of 50%. We therefore generated null distributions by randomly relabeling each participant's data with 50% probability and averaging the data across participants. This step was repeated 10,000 times, yielding a 10,000-element null distribution for each time point. Finally, we implemented a cluster-based permutation test with cluster-forming and cluster size thresholds of $p < .05$ (two-tailed) to evaluate observed differences with respect to the null distribution while accounting for signal autocorrelation.

Statistical analyses of epoch-based decoding accuracy (e.g., Figure 2C-D, Figure 4D, and Figure 5E) were based on bootstrap tests. We simulated population-level decoding accuracy by randomly selecting (with replacement) and averaging data from N of N volunteers (where N is the total number of participants in each experiment) 10,000 times. We then calculated the proportion of permutations where each sample mean was less than or equal to chance decoding accuracy (50%) or 0 (e.g., for analyses comparing decoding accuracy during correct vs. all trials). We also used these distributions to calculate 95% confidence intervals for each statistic; these are the error bars shown in the figures below.

Data & Code Availability. Raw data, stimulus presentation software, and analytic software sufficient to produce each manuscript figure will be made publicly available upon publication of this paper.

<u>Experiment 1</u>.

Data Exclusion. Data from 6 of the 34 volunteers who completed Experiment 1 were discarded. One volunteer withdrew from the study after completing only 2 experimental

blocks. Data from 5 other volunteers were discarded due to hardware or software faults (e.g., codes marking trial events of interest were not sent from the stimulus presentation computer to the eyetracker recording computer, or monocular instead of binocular recordings were obtained). Thus, the data reported here reflect the remaining 28 volunteers.

Experimental Task. A trial schematic is shown in Figure 1A-B. We defined a set of oriented stimuli ±2°, ±5°, ±15°, and ±45° relative to the participant's category boundary, with negative values assigned to Category 1 and positive values assigned to Category 2. Participants reported the category membership of presented stimuli by pressing the "z" and "?" keys (for Categories 1 and 2, respectively) at any point during each three-second trial. Failures to respond within the 3 second trial window were treated as incorrect. We also included trials where the presented stimulus precisely matched the category boundary, i.e., ±0°. Participants were not informed of this manipulation. To maintain the ruse, participants were told to guess if they were uncertain, and random feedback (i.e., correct vs. incorrect) was given at the end of these trials. Participants completed 9 (N = 1), 11 (N = 1), 16 (N = 1), 17 (N = 1), and 18 (N = 24) blocks of 36 trials.

Experiment 2.

Data Exclusion. Data from 2 of the 27 volunteers who completed Experiment 2 were discarded. One volunteer withdrew from the study after completing only 1 experimental block, and data from a second volunteer were discarded due to a hardware fault that occurred midway through testing. Thus, the data reported here reflect the remaining 25 volunteers.

Experimental Task. A task schematic is shown in Figure 3A-B. Participants viewed successive displays of sample and test stimuli separated by a blank interval. We defined a set of oriented stimuli ±15°, ±30°, ±45°, ±60° and ±75° relative to the participant's category boundary, with negative values assigned to Category 1 and positive values assigned to Category 2. Participants were required to report whether the sample and test stimuli were drawn from the same category, pressing "z" to indicate "yes" and "?" to indicate "no". Participants could respond at any point during a 3 second window beginning with the test display, and failures to respond within this window were treated as incorrect. The sample and test stimuli could be drawn from either category, with the exception that the same stimulus value (e.g., +15°) could not be presented as both the sample and the test. Unlike Experiment 1, this task allowed us to decouple categorization from response preparation: since participants cannot select or plan a response before the appearance of the test stimulus, any category-selective signals measured during the sample and delay periods cannot be ascribed to motor processes. Participants completed 9 (N = 2), 13 (N = 1), 14 (N = 1), 17 (N = 1) and 18 (N = 20) blocks of 40 trials.

Experiment 3.

Data Exclusion. Data from 3 of the 16 volunteers who completed Experiment 3 were discarded. Three volunteers withdrew from the study, and data from two volunteers were discarded due to hardware faults that occurred mid-testing. Thus, the data reported here reflect the remaining 13 volunteers.

Experimental Tasks. Participants performed two tasks that used identical stimuli (Figure 5A-B). Participants viewed displays containing flickering apertures of iso-oriented black and white bars. The proportion of white to dark bars varied across trials from 55% to 85% in 10% steps.

During the first half of the experiment, participants performed a categorization task that was nearly identical to Experiment 1. We defined a set of oriented stimuli ±2.5°, ±5.0°, ±12.5°, and ±30° relative to the participant's category boundary, with negative values assigned to Category 1 and positive values assigned to Category 2. Participants reported the category membership of presented stimuli by pressing the "z" and "?" keys (for Categories 1 and 2, respectively) at any point during each three-second trial. Failures to respond within the 3 second stimulus window were treated as incorrect. Participants completed 7 (N = 8) or 8 (N = 5) blocks of 64 trials.

During the second half of the experiment, participants viewed physically identical stimuli but instead reported whether they contained a greater proportion of white vs. black bars regardless of orientation category membership, pressing "z" to indicate a greater proportion of white bars and "?" to indicate a greater proportion of black bars. Failures to respond within the 3 second stimulus window were treated as incorrect. Participants completed 4 (N = 3), 6 (N = 1), 7 (N = 6), or 8 (N = 3) blocks of 64 trials.

Results

Training Tasks

All participants completed a training regimen before starting the main experimental tasks. For each participant and each experiment, we randomly selected a category boundary on the interval (0°,179°] and defined stimuli with orientations ±15°, ±30°, ±45°, and ±60° (Experiments 1 and 3; Figure S1A-B) or ±15°, ±30°, ±45°, ±60° and ±75° relative to the boundary (Experiment 2). Stimuli with negative values (i.e., counterclockwise from the boundary) were assigned to Category 1, while stimuli with positive values (i.e., clockwise from the boundary) were assigned to Category 2. Participants learned to categorize stimuli through trial-and-error, with feedback given at the end of each trial. Participants learned to categorize stimuli quickly, frequently reaching asymptotic performance in 1-2 blocks of 32 trials (Figure S1C-E). Each participant's unique category boundary was used in subsequent experimental tasks.

Experiment 1

Participants in Experiment 1 classified oriented stimuli into discrete groups based on a learned boundary (Figure 1A). To make the categorization task more difficult, and to ensure that participants classified stimuli by rule vs. rote memory, we defined a stimulus set containing some orientation values participants had not seen during training (±0, ±2°, ±5°, ±15°, ±45°; Figure 1B). As expected, participants' accuracy increased, and their response times decreased, with angular distance from the category boundary. We found no differences in behavioral performance between stimuli tilted counterclockwise vs. clockwise relative to the category boundary, therefore, we labeled conditions according to their absolute angular distance from the boundary (Figure 1C).

Eye Movements Reflect Abstract Categories.

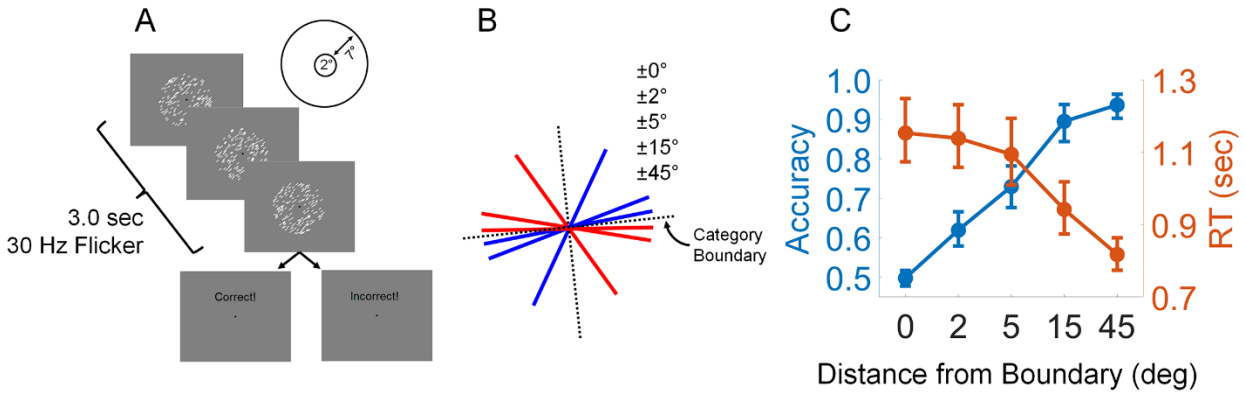


Figure 1. Overview of Experiment 1. *(A) Volunteers viewed displays containing circular apertures of iso-oriented bars. (B) Participants classified stimulus orientations into discrete groups based on an arbitrary (experimenter-imposed) category boundary. Participants could respond at any point during the 3 second stimulus window, and feedback (correct, incorrect) was given at the end of each trial. Note that this schematic is for illustrative purposes and does not contain all possible stimulus orientations (i.e., ±0° and ±2°) (C) Participants' task performance scaled inversely with the angular difference between presented orientations and the category-defining boundary. Error bars depict 95% confidence intervals of the mean.*

We tested whether participants' eye movements varied with stimulus meaning by training a linear classifier to predict stimulus categories from time-resolved patterns of cartesian gaze position. This analysis omitted data from ambiguous trials (i.e., ±0°); we return to these trials below. Eye position data from N-1 experimental blocks were used to train the classifier, and the trained classifier was used to predict category membership from data in the held-out block. This procedure was repeated until every experimental block had served as the test data set. This analysis was performed on a

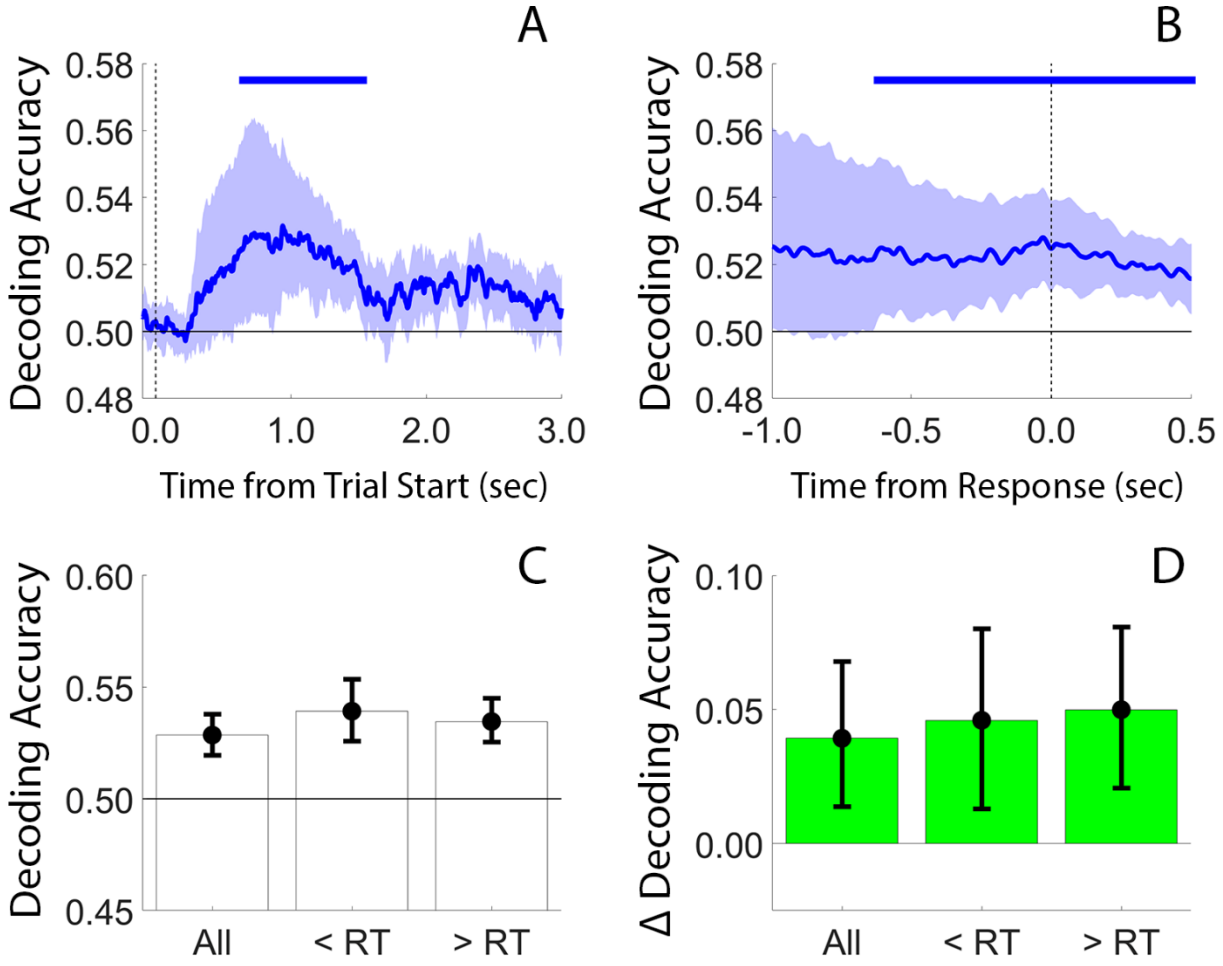


Figure 2. Successful Decoding of Category Membership from Gaze Position. *(A) We trained a linear discriminant classifier to predict stimulus category (i.e., Category 1 or Category 2) from participants gaze data. Decoding accuracy was modestly but robustly above chance during a period spanning ~0.8-1.1 sec after trial start, when most responses occurred (see Figure 1C). (B) Category decoding accuracy remained constant over a period spanning -1.0 to +0.5 sec around participants' responses. (C) Category decoding accuracy calculated from stable fixations during different trial epochs. "All" refers to fixations measured at any point during the 3.0 second stimulus interval, "< RT" refers to fixations measured at any point before a response time threshold, defined as the 80th percentile of each participant's response time distribution calculated over all trials and conditions), and "> RT" refers to fixations measured at any point after the response time threshold. (D) Differences in category decoding accuracy for classifiers trained on fixations measured during correct trials vs. fixations measured during all trials, i.e., correct trials minus all trials. Shaded areas and error bars in each panel depict the 95% confidence interval of the mean. The horizontal lines at 0.5 depict chance decoding accuracy. The horizontal bars at the top of panels A and B depict epochs where category decoding accuracy exceeded chance levels (nonparametric signed randomization tests; Methods).*

timepoint-by-timepoint basis, yielding a single decoding time course per participant. Since there are only two categories, observed decoding accuracy was evaluated relative to a chance value of 50%. This analysis revealed accuracy but robust above-chance decoding of category membership, mostly within an interval spanning ~0.7 to ~1.4 seconds after stimulus onset (Figure 2A). Note that this interval corresponds to the period during which most behavioral responses occurred (Figure 1C). To better visualize relationships between decoding accuracy and behavioral responses, we re-computed decoding accuracy over an interval from -1.0 to +0.5 sec around participants' responses. Decoding accuracy was stable during this period, but variability in decoding accuracy decreased near the time of a behavioral response (Figure 2B).

Because time-resolved category decoding accuracy could be jointly influenced by fixations and saccades, we re-computed category decoding accuracy using only stable fixations (i.e., saccade endpoints). Specifically, we trained linear support vector machines to classify stimulus category based on the cartesian coordinates of stable fixations measured at any point during each 3 second trial ("all"), those measured before a response time threshold, defined as the 80th percentile of each participant's response time distribution across all trials and conditions ("< RT"), and those measured after the response time threshold ("> RT"). Fixations were identified using a velocity-based threshold procedure (see Methods), and we discarded fixations that fell outside the 9 DVA stimulus aperture. This analysis revealed robust decoding of stimulus category (Figure 2C; $p < 1e\text{-}04$ for the "all", "< RT", and "> RT" epochs; bootstrap tests). Identical results were obtained when we restricted this analysis to fixations inside the inner radius of the stimulus aperture (2 DVA; Figure S2A).

We also asked whether stable fixations reflect a decision variable that describes participants' category judgments. This variable is correlated with but different from the presented categories since participants make errors, particularly when the presented stimulus is adjacent to the category boundary. We reasoned that if stable fixations index a decision variable, then category decoding accuracy should be higher during correct vs. incorrect trials. Although our data set contained too few incorrect trials to enable direct comparisons, we did have sufficient data to undertake a comparison of category decoding accuracy using classifiers trained using all trials vs. only correct trials. This analysis revealed modestly but significantly greater category decoding accuracy during correct trials (Figure 2D; $p = 4e\text{-}04$, $p = 0.005$, and $p = 1e\text{-}04$ for the "all", "< RT", and "> RT" epochs, respectively; bootstrap tests), indicating that category-level differences in eye movements reflected an abstract decision variable guiding participants' judgments. We observed similar, though less robust results when we trained classifiers on stable fixations within the inner radius of the stimulus aperture (2 DVA, respectively; Figure S2B).

Category Decoding Depends on Combinations of Eye Movement Frequency, Amplitude, and Direction.

Next, we sought to determine the eye movement parameter(s) responsible for above-chance category decoding accuracy. Visual inspection of participants' fixations revealed no obvious category-specific patterns (Figure 3A). We first tested whether decoding accuracy was driven by differences in the number of saccades recorded during Category 1 or Category 2 trials (irrespective of saccade amplitude or direction). Although saccade frequencies varied substantially across participants (e.g., from a few hundred to a few thousand), we found no evidence for category level-differences in saccade frequency (Figure 3B). Similarly, total saccade count was uncorrelated with category decoding accuracy (Figure S3A). Next, we considered whether decoding was driven by category-level differences in saccade size (irrespective of frequency and direction). Most saccades were small (e.g., within 2-3 DVA of fixation) and their amplitudes were remarkably similar across categories (Figure 3C). Similarly, individual differences in average saccade size were uncorrelated with category decoding accuracy (Figure S3B). Finally, we tested whether decoding

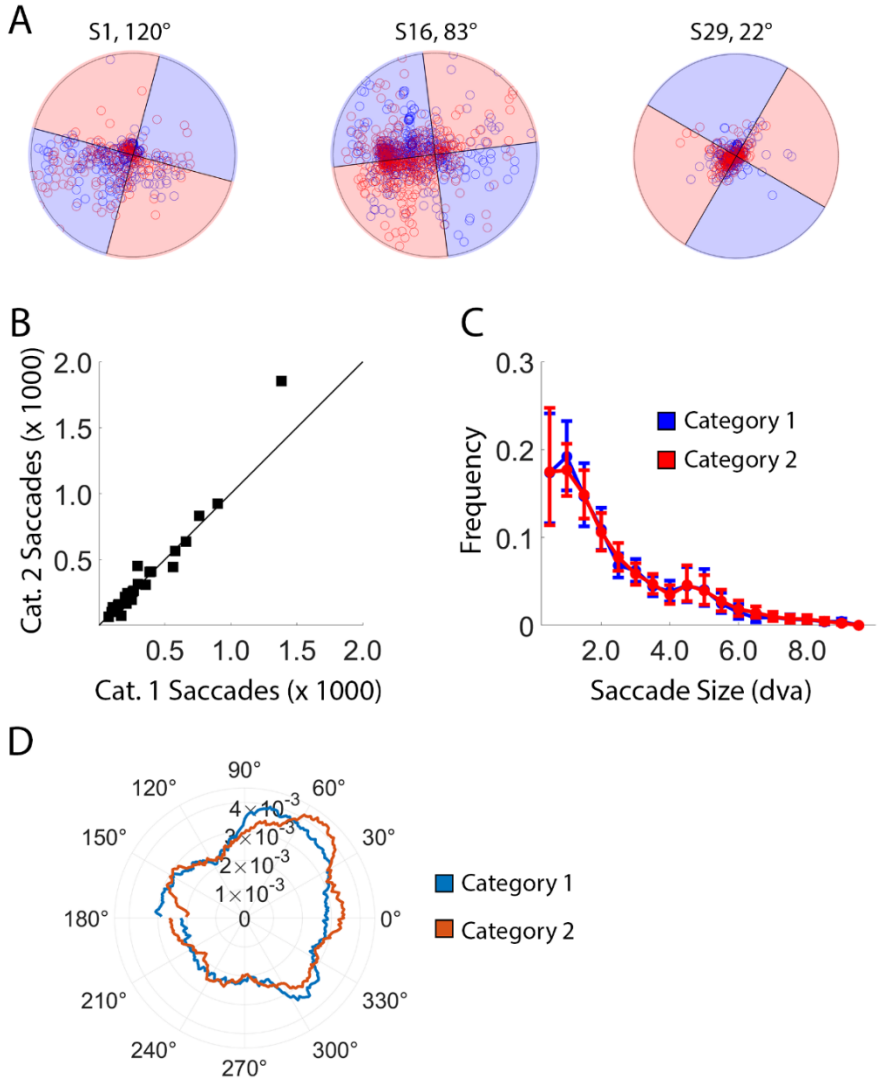


Figure 3. Oculomotor Behavior During Experiment 1. *(A) Plots of individual fixations recorded across the entire experimental testing session for three representative participants with category boundaries of 120°, 83°, and 22°, respectively. Blue and red circles reflect fixations during Category 1 and Category 2 trials, while the shaded areas in each plot depict the range of orientations assigned to Categories 1 and 2 for each participant. The outer radius of each plot corresponds to the outer radius of the stimulus aperture, i.e., 9 DVA. (B) Total fixation counts during Category 1 and Category 2 trials, plotted against a unity line. Each square is an individual participant. (C) Histogram of fixation amplitudes. Most fixations were small, within ~2-3° of the display center. (D) Polar angles of fixations plotted as a function of category membership. Eye position data from each participant were aligned to a common boundary of 0°; thus, in this plot values from 0° to 90° and 180° to 270 degrees correspond to Category 1 orientations, while values from 90° to 180° and 270° to 359° correspond to Category 2 orientations.*

accuracy was driven by category-level differences in saccade direction. For this analysis, we aligned participants fixation records to a common orientation (0°; note that this step was necessary because each participant classified stimuli according to a bespoke category boundary) and constructed circular histograms plotting gaze frequency across each of 360 bins (i.e., 1° bin widths). Saccade directions were remarkably similar and statistically uniform during both Category 1 and Category 2 trials (Figure 3D). Thus, although linear classifiers could successfully decode category membership from records of participants' eye movements, we could not identify a single eye movement parameter responsible for these differences. We suspect that category decoding is driven by small, phasic, and idiosyncratic changes in mixtures of saccade frequencies, amplitudes, and directions. This possibility is supported by a recent non-human primate study, which reported modest, transient, and inconsistent relationships between eye movement parameters and category membership (Rosen & Freedman, 2025). We return to this point in the discussion.

Finally, we note that Experiment 1 confounded category membership with behavioral outputs: volunteers pressed one button if a stimulus was a member of Category 1 and another button if the stimulus was a member of Category 2. We reasoned that if category decoding was driven by response selection or manual actions, then it should be possible to decode participants' responses on ambiguous trials (i.e., when the presented orientation precisely matched the participants' category boundary) from a classifier trained on data from unambiguous trials (i.e., ±2°, ±5°, ±15°, or ±45°). This analysis failed (Figure S4), suggesting that category decoding accuracy did not result from eye movements related to action planning. We expand on this issue in Experiment 2.

Experiment 2

As noted above, the design of Experiment 1 yoked category membership to motor outputs, raising the possibility that category decoding accuracy was driven by factors related to the generation or execution of a manual (button press) response. Experiment 2 addressed this possibility using a delayed match-to-category task (DMC; Freedman & Assad, 2006). Participants viewed displays containing two successive stimuli separated by a blank delay and judged whether the two stimuli were drawn from the same category (Figure 4A). This task requires participants to categorize the first stimulus (the sample) and remember this information over the subsequent delay. Since participants cannot plan a response before presentation of the second stimulus (the test), any above-chance decoding of category membership during the sample and delay periods cannot be ascribed to decision making or motor processing.

Participants performed the DMC task well, and we found no differences in behavioral accuracy during same and different category trials (Figure 4B). As in Experiment 1, we trained a linear discriminant classifier to distinguish between sample categories (i.e., Category 1 or Category 2) using time-resolved records of participants'

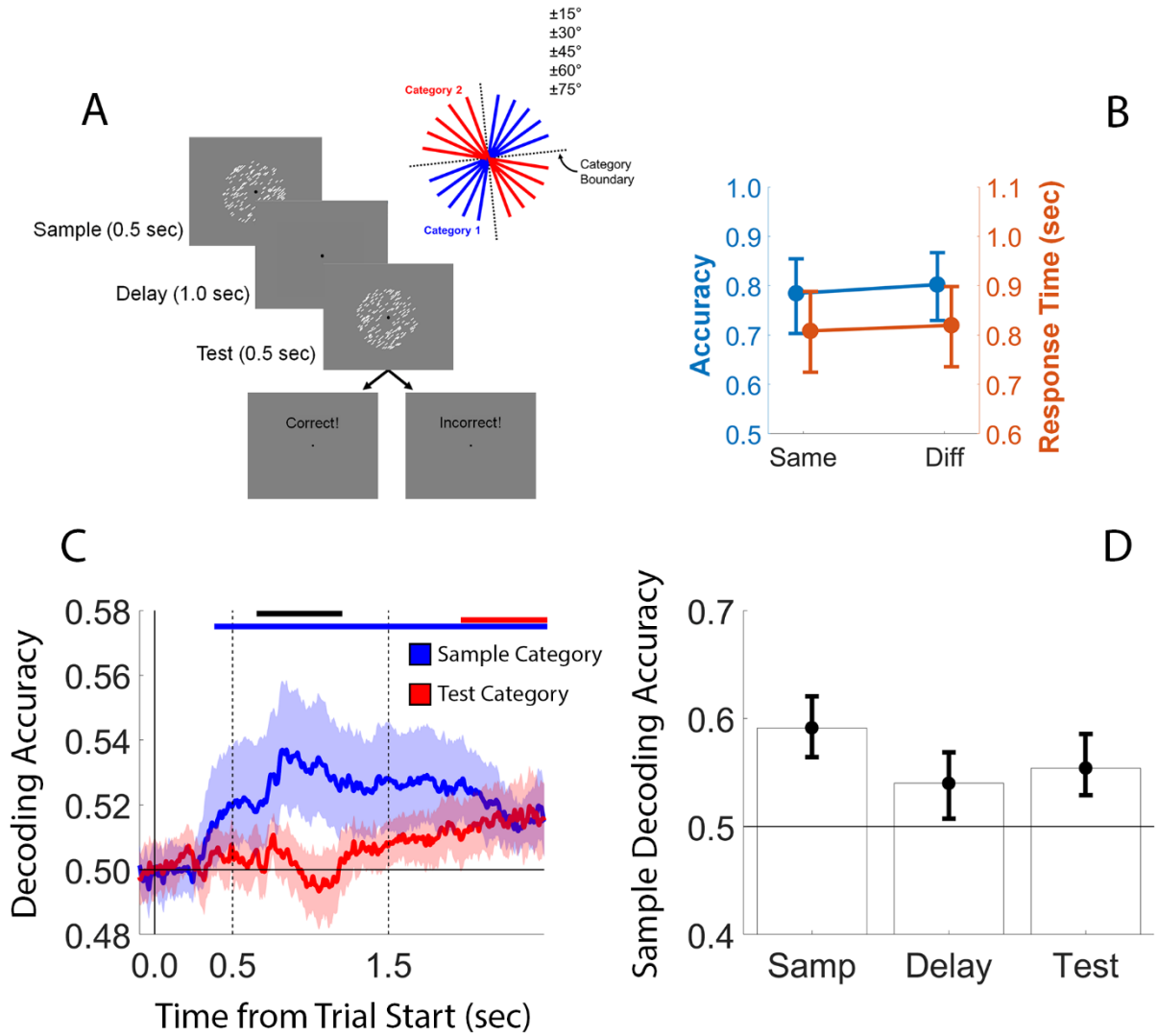


Figure 4. Results of Experiment 2. *(A) Participants performed a delayed match-to-category (DMC) task where they judged whether two successively presented stimuli (the sample and probe) were drawn from the same category. (B) Behavioral performance during same and different trials. (C) A linear discriminant function trained to classify the sample stimulus' category supported robust above-chance decoding accuracy ~0.4 sec after sample onset and throughout the subsequent delay and test periods. A separate decoder trained to classify the test stimulus' category also supported robust above-chance decoding accuracy ~0.4 sec after test onset. The solid vertical line at time 0.0 depicts the onset of the sample stimulus, while the dashed vertical lines at time 0.5 and 1.5 depict the onset of the delay and test periods. (D) Support vector machine classifiers trained on stable fixations measured during the sample, delay, and test periods supported above-chance decoding of the sample category. Shaded regions and error bars in panels B-D depict the 95% confidence interval of the mean; the horizontal solid line at 0.5 in panels C and D depicts chance decoding accuracy. The blue bar and red bars at the top of panel C depicts epochs where sample and test category decoding accuracy was significantly above chance (nonparametric signed randomization test; see Methods). The black bar depicts epochs where sample decoding accuracy was significantly higher than test decoding accuracy.*

gaze behavior. The trained classifier supported modest but statistically robust category decoding accuracy beginning ~400 ms after the appearance of the sample stimulus and throughout the subsequent delay and test periods (Figure 4C). Classifiers trained to

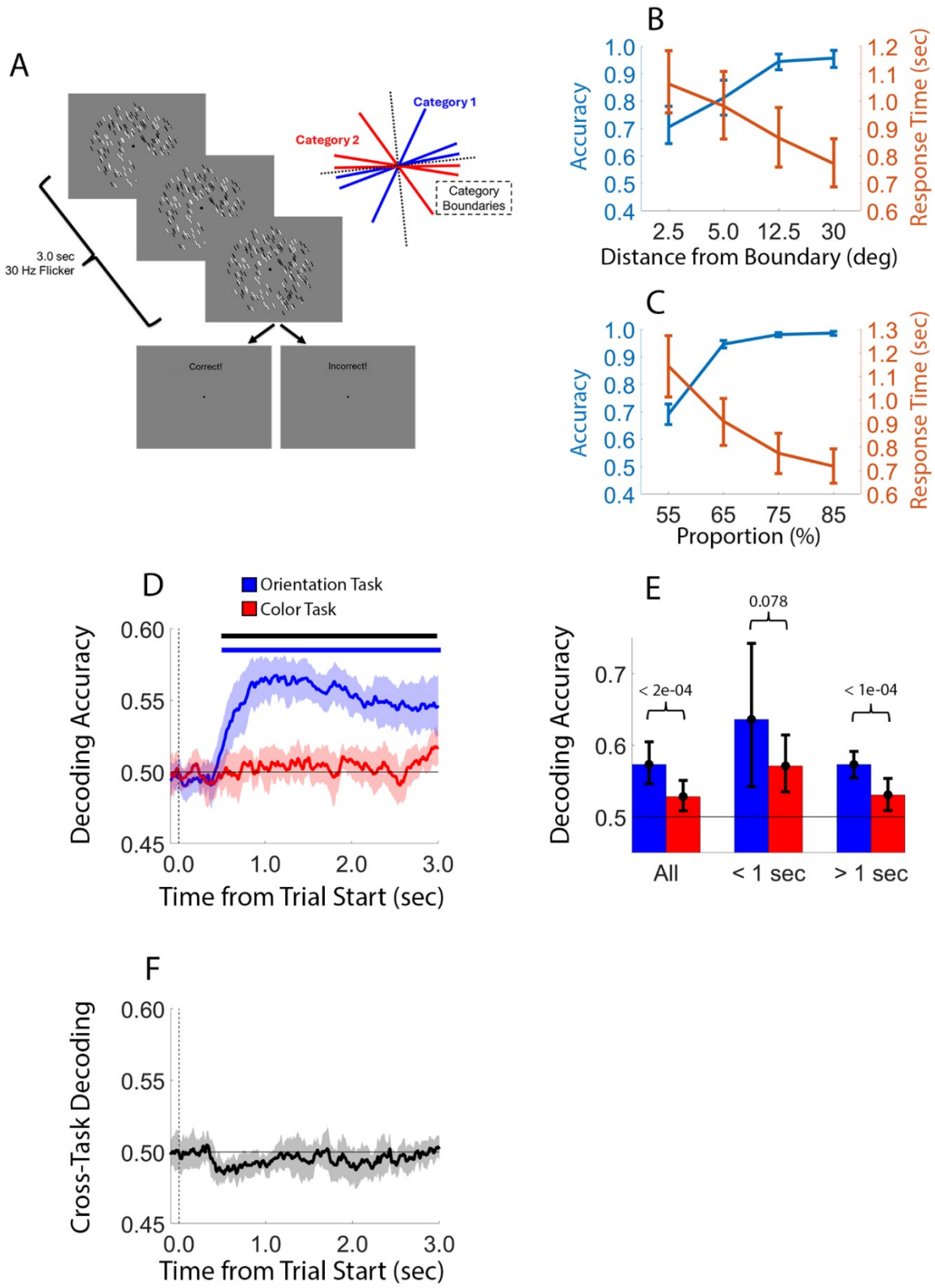
decode the sample category from stable fixations measured during the sample, delay, and test periods also supported above-chance category decoding accuracy (Figure 4D; $p < 1e-04$, $p = 0.008$, and $p < 1e-04$ for the sample, delay, and test epochs, respectively; bootstrap tests). Unlike Experiment 1, category decoding accuracy was equivalent for classifiers trained on fixations from correct trials vs. all trials (Figure S5). Nevertheless, these data provide converging evidence that category decoding accuracy cannot be explained by eye movements linked to motor strategies.

Experiment 3

The designs of Experiments 1 and 2 confound orientation values with category labels. That is, different orientations are assigned to different categories, and thus, what we have described as category-specific gaze patterns could instead be orientation-specific gaze patterns. Experiment 3 tested this possibility via tasks that used identical stimulus sets but required different judgments. During the first half of experimental blocks participants performed an orientation categorization task identical to Experiment 1 (orientation task; Figure 5A). During the second half of experimental blocks participants viewed identical stimuli but instead judged whether they contained a greater proportion of black or white bars (color task; Figure 5A). The latter task requires participants to attentively process the stimulus but does not require an orientation category judgment. We reasoned that if the results of Experiments 1 and 2 reflect processes related to categorization it should be possible to decode category membership from patterns of eye movements during the orientation but not the color task. Obversely, if these results reflect orientation-specific gaze patterns it should be possible to decode “category membership” during both the orientation and color tasks, since they use a common set of oriented stimuli.

Behavioral performance during the orientation and color tasks were similar (Figures 5B and 5C, respectively), though accuracy was generally higher during the color task. Note that, due to a coding error, response time data during the orientation and color tasks for one participant were irretrievably lost; thus, the data plotted here reflect 12 participants instead of 13. Critically, orientation category could be decoded from continuous records of participants’ gaze position during the orientation task, but not during the color task (Figure 5D). Analyses of decoding accuracy computed from stable fixations revealed a more nuanced picture. Note that, because response time data from one participant were lost, we calculated decoding accuracy from fixations measured at any point during the 3 second trial, during the first second of each trial (i.e., when most responses occurred), and during the last two seconds of each trial (i.e., after most responses were complete). Although it was possible to decode orientation category during different epochs of the color task ($p = 0.003$, $p < 1e-04$, and $p = 0.003$ for fixations measured at any point during the 3 second trial, during the first second of each trial, or during the last two seconds of each trial, respectively; bootstrap tests), decoding accuracy was generally greater during the orientation compared to the color task ($p = 2e-04$, $p = 0.078$, and $p < 1e-04$ for the “all”, “< 1 sec”, and “> 1 sec” conditions, respectively; Figure 5E).

Finally, we implemented cross-task decoding analyses to test the extent to which the orientation and color tasks evoked similar patterns of eye movements. In the first



phase of the analysis, we trained a linear discriminant classifier to predict stimulus category using records of gaze position measured during the orientation task. The

Fig 5. Results of Experiment 3. *(A) Participants performed two tasks while viewing displays of circular apertures containing black and white bars. During the first half of the experiment, participants classified stimulus orientation according to a memorized boundary (orientation task; inset). During the second half of the experiment, participants classified stimuli according to whether they contained a greater proportion of white or black bars (color task). (B) Performance during the orientation task. (C) Performance during the color task. (D) Orientation category decoding accuracy during the orientation and color tasks. The blue and black bars at the top of the plot mark epochs where orientation category decoding accuracy was greater than chance and when orientation decoding accuracy was greater during the orientation task relative to the color task. Support vector machine classifiers trained on stable fixations measured during the during the entire stimulus epoch ("all") the first second of each trial ("< 1 sec") and the last two seconds of each trial ("> 1 sec") supported above-chance category decoding during the orientation and color tasks. However, decoding accuracy was generally higher during the orientation compared to the color task. (F) Cross-task decoding analysis. A classifier was trained to predict stimulus category using data from the orientation task, and the trained classifier was used to predict stimulus category using data from the color task. This analysis failed to reveal above-chance decoding accuracy at any point during the 3 second stimulus window. Thus, the orientation and color tasks evoked different patterns of eye movements, even though they used a common set of oriented stimuli. Shaded regions and error bars in each plot depict the 95% confidence interval of the mean.*

trained classifier was then used to predict stimulus category from records of gaze position during the color task. This analysis failed to reveal above-chance category decoding accuracy at any point during the 3 second stimulus epoch (Figure 5F). Next, we trained a support vector machine to predict stimulus category based on stable fixations measured at any point during the stimulus epoch during the orientation task, then used the trained classifier to predict stimulus category from stable fixations measured at any point during the color task. This analysis also failed to reveal above-chance category decoding (M = 0.482, p = 0.856, bootstrap test). Thus, although it was possible to decode category membership during the color task (e.g., Figure 5E), the patterns of eye movements that supported category decoding were different during the orientation and color tasks. Since the orientation and color tasks used a common set of oriented stimuli, this result suggests that category decoding accuracy cannot be attributed to orientation-specific eye movements.

Discussion

Recent studies have demonstrated that human oculomotor behavior tracks the locations (van Ede et al., 2019) and orientations (Linde-Domingo & Spitzer, 2024) of items held in WM. Our findings build on this work by demonstrating that human oculomotor behavior tracks stimulus meaning, operationalized as category membership. Whereas the DMC task used in Experiment 2 engages WM by requiring participants to remember the category membership of a stimulus over a short delay, we argue that the categorization tasks used in Experiments 1 and 3 engage WM by requiring participants to rapidly compare sensory inputs with an internal representation of a category-defining boundary. Long-term memory (LTM) may also play a role: the volunteers tested in this study classified stimuli according to a boundary that was held constant across a 90-minute testing session, and human EEG studies suggest that target-defining information is rapidly transferred from WM to LTM when task rules are held constant (Carlisle et al., 2011). Similarly, a recent electrocorticography study demonstrated that category-selective neural activity in the human hippocampus is similar during tasks with short- and long memory delays (Daume et al., 2024), suggesting a cellular correspondence between category-selective neural responses during WM and LTM. Further research is needed to explore the roles that brain areas linked with short- and long-term memory play in maintaining category-defining information and how define how these systems interact with brain areas responsible for oculomotor control.

Our findings align with the results of a recent non-human primate study (Rosen & Freedman, 2025) that quantified the oculomotor behavior of several monkeys trained to perform variants of DMC tasks and that found strong evidence for category-specific eye movements. A major strength of this work is that gaze position records were obtained concurrently with neural recordings in oculomotor brain areas, including SC and LIP. Analyses of neural activity in these areas suggested that representations of stimulus category and gaze position were represented by orthogonal subspaces of population-level activity. That is, patterns of category-selective and eye movement-selective neural activity were represented using a coding scheme designed to minimize cross-talk between these signals. The authors of this study (see also Peysakovich et al., 2024) speculated that category-selective eye movements during the DMC task are a result of “leakage” between neural subspaces representing category and eye position information. We suspect that a similar process is responsible for the category-specific eye movements we report here, though additional research will be needed to test this possibility.

Although human gaze position recordings supported above-chance category decoding accuracy (Figures 2, 4, and 5), we could not identify a single parameter – i.e., eye movement frequency, amplitude, or direction – that was sufficient to explain this result (Figure 3). This outcome aligns with findings reported by Rosen and Freedman (2025), who found modest, inconsistent, transient, and idiosyncratic links between eye movement parameters and category membership. Specifically, these authors found that different eye movement parameters (e.g., direction, size, frequency) carried category information at different times and in different animals, and that these effects rarely lasted for longer than a few hundred milliseconds. Collectively, those findings and the results of the current study suggest that category-specific gaze patterns can arise

through transient differences in a variety of eye movement parameters invisible to the naked eye.

There are several salient differences between this study and the experiments surveyed by Rosen and Freedman (2025). First, many of the non-human primate studies reviewed by Rosen and Freedman presented stimuli extrafoveally. This choice was born of necessity: since these studies were designed to identify category-selective neural signals in various brain areas (e.g., SC, LIP) it was necessary to present stimuli in the spatial receptive fields of target neurons. Nevertheless, this raises the possibility that the category-specific patterns of eye movements reported by Rosen and Freedman were jointly influenced by stimulus location (i.e., a spatial code) and stimulus category. Our study suggests that this is not the case, since we observed category-specific differences in oculomotor behavior for stimuli presented at fovea. Second, the monkeys tested in the experiments surveyed by Rosen and Freedman completed tens- to hundreds-of-thousands of training trials spread over several months, raising the possibility that category-specific eye movements are a by-product of hypertraining. Again, our data indicate that this is not the case: with the benefit of verbal instructions, our human volunteers reached asymptotic categorization performance in ~100 trials (~8-10 minutes), and we observed category-specific differences in oculomotor behaviors that were broadly consistent with those reported by Rosen and Freedman (2025). We do not suggest that training regimens and expertise have no influence on oculomotor behavior. After all, whereas gaze-position-based category decoding reaches ~75% in monkeys (Rosen & Freedman, 2025), decoding accuracy in the current study ranged from ~53-62%.

The rule-based tasks used in this study (and many nonhuman primate studies) reflect only one form of categorization. Humans and other animals are adept at learning novel categories defined by combinations of different features, for example, orientation and spatial frequency (e.g., Goldstone, 1994; Reinert et al., 2021; Henderson et al., 2025). Whereas rule-based categorization tasks tend to recruit brain areas associated with WM (e.g., the parietal and prefrontal cortex), these so-called information integration tasks recruit neural circuits associated with procedural learning in the striatum and basal ganglia (Ashby & Maddox, 2005). On the one hand, information integration tasks may evoke category-specific eye movement patterns like those seen during rule-based tasks via connections between the basal ganglia and SC through the subthalamic nucleus (Melleu & Sabino Canteras, 2024). On the other hand, information integration tasks may not evoke category-selective eye movements due to recruitment of fundamentally different neural circuits. Notably, a recent human behavioral study reported that rule-based category learning generalized across visual hemifields while information-integration-based category learning did not (Rosedahl et al., 2018; see also Tan et al., 2019).

Information stored in WM guides what we need to do next: where to look, where to grasp, or what to say. Whereas prior studies have reported stimulus- and location-specific patterns of eye movements relative to items held in WM, our findings reveal a behavioral “tell”, where stimulus meaning can be extracted from subtle variations in where humans look. We regard this as an important conceptual advance, since in naturalistic settings our actions are determined by stimulus meaning. Similarly, our findings extend prior non-human primate research by demonstrating that category-

specific eye movements are not a by-product of extrafoveal stimulus presentation or extensive training regimens. Thus, our findings provide converging evidence linking systems for WM and oculomotor control and suggest that eye movements encode both physical and semantic variables.

References

Ashby, F. G., & Maddox, W. T. (2005). Human Category Learning. *Annual Review of Psychology*, *56*(1), 149–178. https://doi.org/10.1146/annurev.psych.56.091103.070217

Carlisle, N. B., Arita, J. T., Pardo, D., & Woodman, G. F. (2011). Attentional Templates in Visual Working Memory. *Journal of Neuroscience*, *31*(25), 9315–9322. https://doi.org/10.1523/JNEUROSCI.1097-11.2011

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, *24*(1), 87–114. https://doi.org/10.1017/S0140525X01003922

Ester, E. F., Anderson, D. E., Serences, J. T., & Awh, E. (2013). A Neural Measure of Precision in Visual Working Memory. *Journal of Cognitive Neuroscience*, *25*(5), 754–761. https://doi.org/10.1162/jocn_a_00357

Ester, E. F., Serences, J. T., & Awh, E. (2009). Spatially Global Representations in Human Primary Visual Cortex during Working Memory Maintenance. *The Journal of Neuroscience*, *29*(48), 15258–15265. https://doi.org/10.1523/JNEUROSCI.4388-09.2009

Ester, E. F., Sprague, T. C., & Serences, J. T. (2015). Parietal and Frontal Cortex Encode Stimulus-Specific Mnemonic Representations during Visual Working Memory. *Neuron*, *87*(4), 893–905. https://doi.org/10.1016/j.neuron.2015.07.013

Ester, E. F., Sprague, T. C., & Serences, J. T. (2020). Categorical Biases in Human Occipitoparietal Cortex. *The Journal of Neuroscience*, *40*(4), 917–931. https://doi.org/10.1523/JNEUROSCI.2700-19.2019

Ester, E., & Weese, R. (2023). Temporally Dissociable Mechanisms of Spatial, Feature, and Motor Selection during Working Memory–guided Behavior. *Journal of Cognitive Neuroscience*, *35*(12), 2014–2027. https://doi.org/10.1162/jocn_a_02061

Freedman, D. J., Riesenhuber, M., Poggio, T., & Miller, E. K. (2001). Categorical Representation of Visual Stimuli in the Primate Prefrontal Cortex. *Science*, *291*(5502), 312–316. https://doi.org/10.1126/science.291.5502.312

Funahashi, S., Chafee, M. V., & Goldman-Rakic, P. S. (1993). Prefrontal neuronal activity in rhesus monkeys performing a delayed anti-saccade task. *Nature*, *365*(6448), 753–756. https://doi.org/10.1038/365753a0

Fuster, J. M., & Alexander, G. E. (1971). Neuron Activity Related to Short-Term Memory. *Science*, *173*(3997), 652–654. https://doi.org/10.1126/science.173.3997.652

Gallivan, J. P., Chapman, C. S., Gale, D. J., Flanagan, J. R., & Culham, J. C. (2019). Selective Modulation of Early Visual Cortical Activity by Movement Intention. *Cerebral Cortex*, *29*(11), 4662–4678. https://doi.org/10.1093/cercor/bhy345

Goldstone, R. L. (1994). Influences of categorization on perceptual discrimination. *Journal of Experimental Psychology: General, 123*(2), 178–200. https://doi.org/10.1037/0096-3445.123.2.178

Gutteling, T. P., Petridou, N., Dumoulin, S. O., Harvey, B. M., Aarnoutse, E. J., Kenemans, J. L., & Neggers, S. F. W. (2015). Action Preparation Shapes Processing in Early Visual Cortex. *The Journal of Neuroscience*, *35*(16), 6472–6480. https://doi.org/10.1523/JNEUROSCI.1358-14.2015

Harrison, S. A., & Tong, F. (2009). Decoding reveals the contents of visual working memory in early visual areas. *Nature*, *458*(7238), 632–635. https://doi.org/10.1038/nature07832

Henderson, M. M., Serences, J. T., & Rungratsameetaweemana, N. (2023). *Dynamic categorization rules alter representations in human visual cortex*. https://doi.org/10.1101/2023.09.11.557257

Kwak, Y., & Curtis, C. E. (2022). Unveiling the abstract format of mnemonic representations. *Neuron*, *110*(11), 1822-1828.e5. https://doi.org/10.1016/j.neuron.2022.03.016

Linde-Domingo, J., & Spitzer, B. (2023). Geometry of visuospatial working memory information in miniature gaze patterns. *Nature Human Behaviour*, *8*(2), 336–348. https://doi.org/10.1038/s41562-023-01737-z

Liu, B., Alexopoulou, Z.-S., & Van Ede, F. (2024). Jointly looking to the past and the future in visual working memory. *eLife*, *12*, RP90874. https://doi.org/10.7554/eLife.90874

Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods*, *164*(1), 177–190. https://doi.org/10.1016/j.jneumeth.2007.03.024

Mostert, P., Albers, A. M., Brinkman, L., Todorova, L., Kok, P., & De Lange, F. P. (2018). Eye Movement-Related Confounds in Neural Decoding of Visual Working Memory Representations. *Eneuro*, *5*(4), ENEURO.0401-17.2018. https://doi.org/10.1523/ENEURO.0401-17.2018

Peysakhovich, B., Zhu, O., Tetrick, S. M., Shirhatti, V., Silva, A. A., Li, S., Ibos, G., Rosen, M. C., Johnston, W. J., & Freedman, D. J. (2024). Primate superior colliculus is causally engaged in abstract higher-order cognition. *Nature Neuroscience*. https://doi.org/10.1038/s41593-024-01744-x

Quax, S. C., Dijkstra, N., Van Staveren, M. J., Bosch, S. E., & Van Gerven, M. A. J. (2019). Eye movements explain decodability during perception and cued attention in MEG. *NeuroImage*, *195*, 444–453. https://doi.org/10.1016/j.neuroimage.2019.03.069

Reinert, S., Hübener, M., Bonhoeffer, T., & Goltstein, P. M. (2021). Mouse prefrontal cortex represents learned rules for categorization. *Nature*, *593*(7859), 411–417. https://doi.org/10.1038/s41586-021-03452-z

Serences, J. T., Ester, E. F., Vogel, E. K., & Awh, E. (2009). Stimulus-Specific Delay Activity in Human Primary Visual Cortex. *Psychological Science*, *20*(2), 207–214. https://doi.org/10.1111/j.1467-9280.2009.02276.x

Shao, Z., Zhang, M., & Yu, Q. (2025). Stimulus representation in human frontal cortex supports flexible control in working memory. *eLife*, *13*, RP100287. https://doi.org/10.7554/eLife.100287

Snyder, L. H., Batista, A. P., & Andersen, R. A. (1997). Coding of intention in the posterior parietal cortex. *Nature*, *386*(6621), 167–170. https://doi.org/10.1038/386167a0

Sprague, T. C., Ester, E. F., & Serences, J. T. (2014). Reconstructions of Information in Visual Spatial Working Memory Degrade with Memory Load. *Current Biology*, *24*(18), 2174–2180. https://doi.org/10.1016/j.cub.2014.07.066

Sprague, T. C., Ester, E. F., & Serences, J. T. (2016). Restoring Latent Visual Working Memory Representations in Human Cortex. *Neuron*, *91*(3), 694–707. https://doi.org/10.1016/j.neuron.2016.07.006

Svoboda, K., & Li, N. (2018). Neural mechanisms of movement planning: Motor cortex and beyond. *Current Opinion in Neurobiology*, *49*, 33–41. https://doi.org/10.1016/j.conb.2017.10.023

Tan, Q., Wang, Z., Sasaki, Y., & Watanabe, T. (2019). Category-Induced Transfer of Visual Perceptual Learning. *Current Biology*, *29*(8), 1374-1378.e3. https://doi.org/10.1016/j.cub.2019.03.003

Van Ede, F., Chekroud, S. R., & Nobre, A. C. (2019). Human gaze tracks attentional focusing in memorized visual space. *Nature Human Behaviour*, *3*(5), 462–470. https://doi.org/10.1038/s41562-019-0549-y

Van Ede, F., Chekroud, S. R., Stokes, M. G., & Nobre, A. C. (2019). Concurrent visual and motor selection during visual working memory guided action. *Nature Neuroscience*, *22*(3), 477–483. https://doi.org/10.1038/s41593-018-0335-6

Zhou, Y., & Freedman, D. J. (2019). Posterior parietal cortex plays a causal role in perceptual and categorical decisions. *Science*, *365*(6449), 180–185. https://doi.org/10.1126/science.aaw8347

Zhou, Y., Zhu, O., & Freedman, D. J. (2023). Posterior Parietal Cortex Plays a Causal Role in Abstract Memory-Based Visual Categorical Decisions. *The Journal of Neuroscience*, *43*(23), 4315–4328. https://doi.org/10.1523/JNEUROSCI.2241-22.2023